The cut-sky cosmic microwave background is not anomalous

Andrew Pontzen[1,*] and Hiranya V. Peiris[1,2,†]

[1] *Institute of Astronomy and Kavli Institute for Cosmology, University of Cambridge, Cambridge CB3 0HA, U.K.*

[2] *Department of Physics and Astronomy, University College London, London WC1E 6BT, U.K.*

(Dated: May 21, 2010)

The observed angular correlation function of the cosmic microwave background has previously been reported to be anomalous, particularly when measured in regions of the sky uncontaminated by Galactic emission. Recent work by Efstathiou *et al.* presents a Bayesian comparison of isotropic theories, casting doubt on the significance of the purported anomaly. We extend this analysis to all anisotropic Gaussian theories with vanishing mean ($\langle \delta T \rangle = 0$), using the much wider class of models to confirm that the anomaly is not likely to point to new physics. On the other hand if there is any new physics to be gleaned, it results from low-ℓ alignments which will be better quantified by a full-sky statistic.

We also consider quadratic maximum likelihood power spectrum estimators that are constructed assuming isotropy. The underlying assumptions are therefore false if the ensemble is anisotropic. Nonetheless we demonstrate that, for theories compatible with the observed sky, these estimators (while no longer optimal) remain statistically superior to pseudo-C_ℓ power spectrum estimators.

I. INTRODUCTION

Observations of the cosmic microwave background (CMB) by the *Wilkinson Microwave Anisotropy Probe* (WMAP; *e.g.* [1, 2]) are widely interpreted as confirming the standard model of cosmology in which inflation generates a homogeneous and isotropic background and seeds isotropic, nearly scale-free perturbations. Yet a variety of tests suggest that, on large scales, something may be amiss [3–10]. (For a wide-ranging assessment of such anomalies in the 7-yr WMAP data see Ref. [11].) The interpretation of these results is complicated by the *a posteriori* nature of anomaly hunting: any large dataset will contain statistical flukes which, in isolation, can be made to look unacceptable. This is a particularly pernicious problem in the context of large-scale cosmology: with only one sky to observe, frequentist statistics are almost impossible to interpret.

Frequentist results can be made into more concrete Bayesian statements by considering specific alternative CMB theories or classes of theories (see *e.g.* Refs. [12–14]). But a single, fixed dataset can still contribute overwhelming evidence in favour of or against the very same theory, depending on the alternatives against which we are judging (for an elucidation of this point, see Ref. [15], Sec. 5.5). In other words there is no unique way to ascribe significance to departures from the standard theory.

This does not imply we should abandon critical evaluations of WMAP and other data: if we simply accept we have an 'unlikely' realization of our favoured theory, we might miss the opportunity to discover new physics (or instrumental systematics). Thus frequentist results cannot be dismissed out-of-hand; but we would advocate their interpretation as pointers to interesting areas of work, rather than quantifiable death-knells of existing models or theories.

In the present work, we will consider a long-standing debate about the nature of the angular correlation function $\mathcal{C}(\theta)$ of the CMB. The argument is usually phrased in terms of the statistic $\mathcal{S}^{\rm cut}_{1/2}$, which traces the extent to which temperature fluctuations (outside a Galactic mask) are correlated between points separated by 60° or more. For a quantitative definition, see Section II. A number of recent works have attempted to assess the significance of the purportedly anomalous value of $\mathcal{S}^{\rm cut}_{1/2}$, reaching essentially contradictory conclusions. In particular, the frequentist P-value [10] suggests the observed sky is highly anomalous, while a Bayesian analysis of the optimally reconstructed sky by Efstathiou *et al.* suggests the opposite [16]; see also Ref. [11]. However, any Bayesian result pivots crucially on the alternative models considered; the assumptions in Ref. [16] mean that only isotropic models are considered. This is a significant omission, since it leaves open the possibility that suboptimal estimates of $\mathcal{S}_{1/2}$ formed from cut sky data can be reframed as useful measures of anisotropy.

The present work rectifies that omission. The anomaly is analysed from within harmonic space, and then anisotropic theories which make our CMB realization more probable are considered. The $\mathcal{S}^{\rm cut}_{1/2}$ anomaly is found to be uninformative in the following two senses:

1. The trivial maximum likelihood anisotropic Gaussian theory for our observed sky[1] does not lead to substantially better likelihoods for the single statistic $\mathcal{S}^{\rm cut}_{1/2}$;

*Electronic address: apontzen@ast.cam.ac.uk

†Electronic address: h.peiris@ucl.ac.uk

[1] Namely, that with covariance matrix $\mathbf{C} = \mathbf{a}\mathbf{a}^\dagger$ where $\mathbf{a}$ is the observed sky data vector.

2. Theories constructed specifically to maximize the likelihood of $\mathcal{S}_{1/2}^{\rm cut}$ (ignoring the rest of the information on the sky) also yield little gain.

These failures arise from the large variance inherent in using a statistic, such as $\mathcal{S}_{1/2}^{\rm cut}$, which is quartic in the data. Overall, then, the present work reinforces the view that the frequentist 'unlikeliness' of the observed sky must be regarded as a statistical fluke.

Some broader results arise from our study. First, we consider the effect of an anisotropic theory on quadratic maximum likelihood (QML) estimates of the power spectrum. The QML estimators are derived under the (in this context false) assumption of isotropy; despite this, they typically remain superior to pseudo-C_ℓ approaches to power spectrum estimation (Section II, with detail in Appendix A 5). Second, we present an extremely fast method for finding the maximum angular momentum direction of a CMB map (Appendix E). Third, we demonstrate that cut-sky correlation functions can be exactly reproduced from the pseudo-C_ℓ power spectrum (Appendix B). This final result, applicable also for weighted data, has been reported previously [17] but ignored by recent work; to our knowledge no proof appears in the existing literature.

The paper is structured as follows. Section II introduces the necessary background and notation. In Section III we consider, from a harmonic-space perspective, the origin of the low observed $S_{1/2}^{\rm cut}$. Anisotropic, Gaussian theories which reproduce this result are considered in Section IV, and show that even the best conceivable fit to the observed CMB makes no substantial improvement to the $S_{1/2}^{\rm cut}$ likelihood. Finally, the work is summarized in Section V. Further details and discussion are contained in appendices.

II. BACKGROUND AND NOTATION

In this Section we set out the various definitions needed in our work. Let us start by defining the observed temperature correlation function $\mathcal{C}(\theta)$ as

$$\mathcal{C}(\theta) = \overline{T(\hat{\mathbf{n}}_1)T(\hat{\mathbf{n}}_2)}, \tag{1}$$

where the overbar denotes averaging over all observed line-of-sight vector pairs $\hat{\mathbf{n}}_1$, $\hat{\mathbf{n}}_2$ satisfying $\hat{\mathbf{n}}_1 \cdot \hat{\mathbf{n}}_2 = \cos\theta$. We further define

$$\mathcal{C}_\ell \equiv \frac{1}{2\ell+1}\sum_m |a_{\ell m}|^2, \tag{2}$$

where the $a_{\ell m}$'s are the spherical harmonic coefficients of the temperature field on the observed sky. A calligraphic $\mathcal{C}_\ell$ thus denotes the observed power, distinguished from the theoretical variances C_ℓ which we regard as defined by the relation

$$C_\ell = \langle \mathcal{C}_\ell \rangle \tag{3}$$

for isotropic and anisotropic theories alike.

There is an exact relationship between $\mathcal{C}(\theta)$ and $\mathcal{C}_\ell$, namely

$$\mathcal{C}(\theta) = \frac{1}{4\pi}\sum_\ell (2\ell+1)\mathcal{C}_\ell P_\ell(\cos\theta), \tag{4}$$

where P_ℓ are the Legendre polynomials. Throughout this paper in our numerical calculations, we use a finite sum over $2 \le \ell \le 30$; the lower limit discards any contribution from the monopole and dipole, while we verified that the upper limit is high enough for our results to converge.

Result (4) holds regardless of any theoretical constraints (such as isotropy). Thus the information in the correlation function is identical to that in the $\mathcal{C}_\ell$'s observed on the sky. If desired, one can define the theoretical correlation function to be the ensemble average of the sky-observed correlation function, $C(\theta) = \langle \mathcal{C}(\theta) \rangle$.

The purported anomalies relate to the apparent lack of correlations on large angular scales, quantified by

$$\mathcal{S}_{1/2} = \int_{-1}^{1/2} \mathcal{C}(\theta)^2 \sin\theta \, \mathrm{d}\theta. \tag{5}$$

This quantity is a measure of the extent to which the temperature from points separated by 60° or more is correlated. Rather than evaluate (5) directly, it is much faster and numerically more stable to calculate $\mathcal{S}_{1/2}$ from the quadratic form

$$\mathcal{S}_{1/2} = \sum_{\ell\ell'} \mathcal{C}_\ell \mathcal{C}_{\ell'} s_{\ell\ell'}, \tag{6}$$

where, as above, ℓ and ℓ' range from 2 to 30 in our numerical calculations and

$$s_{\ell\ell'} = \int_{-1}^{1/2} P_\ell(x) P_{\ell'}(x) \mathrm{d}x, \tag{7}$$

which may be computed using well-known recursion relations (*e.g.* Appendix C.2 of Ref. [18]; see also Ref. [10]).

If one does not trust information inside a specified mask (for instance due to suspected Galactic contamination), one may calculate the correlation function using only the points outside the mask,

$$\mathcal{C}(\theta)^{\rm cut} \equiv \left.\overline{T(\hat{\mathbf{n}}_1)T(\hat{\mathbf{n}}_2)}\right|_{M(\hat{\mathbf{n}}_1)=M(\hat{\mathbf{n}}_2)=1}, \tag{8}$$

where $M(\hat{\mathbf{n}})$ is a masking function (equal to 0 or 1 in each pixel), so that the angular average denoted by the overbar is over all point pairs (with $\hat{\mathbf{n}}_1 \cdot \hat{\mathbf{n}}_2 = \cos\theta$) which lie outside the mask. This procedure is mathematically identical to calculating the Legendre sum over the pseudo-C_ℓ (PCL) estimates for the power spectrum (which we denote $\hat{C}_\ell^{\rm PCL}$; see Appendix A for a precise definition):

$$\mathcal{C}(\theta)^{\rm cut} = \frac{1}{4\pi}\sum_\ell (2\ell+1)\hat{C}_\ell^{\rm PCL} P_\ell(\cos\theta). \tag{9}$$

While this result has been reported before [17], an explicit proof does not appear to exist in the literature, so we provide one in Appendix B.

It is clear that, if $\hat{C}_\ell$ are any unbiased estimates for C_ℓ, then forming their Legendre sum (4) yields an unbiased estimator for $C(\theta)$. Thus if one wishes to find maximum likelihood estimates for $C(\theta)$ on the full sky from cut sky information, by linearity one simply substitutes the maximum likelihood $\hat{C}_\ell$ estimates in place of the PCL estimates. For our purposes, the estimates provided by the quadratic maximum likelihood (QML) estimator technique [19] are close enough to the exact maximum likelihood to remove the need for any non-linear techniques [20].

There are two somewhat subtle points to be appreciated here. First, reservations have been expressed about the use of the QML estimator, since it uses prior information on the power spectrum and therefore appears to make strong assumptions about the form of the underlying theory. Copi *et al.* [10] express concern about use of the QML estimator in circumstances where one is questioning the validity of the concordance model. In fact, this unease turns out to be unwarranted; one may explicitly show that QML estimates remain superior to PCL estimates – even in cases where the estimates for the covariance matrix are not correct.

Let us outline why this should be so. A full derivation is given in Appendix A 5, in which the QML estimator is written entirely within harmonic space. The estimation procedure can be seen to down-weight high variance modes and up-weight low variance modes before calculating the power spectrum of the weighted cut sky. The resulting power spectrum is then correctly de-weighted and de-convolved[2]. The overall effect is to minimize the cross-talk from the mask-induced mode-coupling. Accordingly the QML estimator will be close to optimal in reconstructing skies from any theory where the power in each ℓ is close to that predicted in concordance models. Since the observed power spectrum is very close to the ΛCDM theoretical prediction, any serious candidate theory must satisfy this criterion. Indeed, only if the true theory has a power spectrum closer to flat than to the concordance model will the PCL estimator typically perform better than the QML. A corollary is that the PCL and QML estimators become identical for a theory with flat power spectrum[3] (equal C_ℓ's).

Efstathiou et al [16] make a different rebuttal of the sensitivity-to-assumptions concern by showing that full-sky maps (albeit band-limited) can be made from the cut sky data (see also Ref. [7]). The difficulty with full-sky reconstructions is that they rapidly become unstable as the sky cut increases (unless the data are strictly band-limited, which is not true of the CMB). Furthermore they introduce a dependence on the underlying theory so that their conceptual benefits over the QML estimator are not clear (although, as shown in Ref. [16], the sensitivity to the assumed covariance may be rather weak). For this reason we will not consider explicit sky reconstructions further in the present work.

However the second subtlety is that one may not, in fact, want to optimally reconstruct the full sky correlation function. If we are interested in using a cut sky correlation function *not* to remove localized contamination, but instead as a distinct quantity in its own right, the efficiency of the QML estimator at reconstructing the full sky becomes a hinderance. Starting instead from definition (8) – the correlation of pixel pairs in a finite region of the sky – gives us a transparent interpretation. Therefore it remains of interest to examine carefully the PCL-derived $C(\theta)^{\rm cut}$, not because the QML estimator technique is in doubt as a way of extracting reliable full-sky information, but because the PCL technique explicitly extracts information which is different.

III. WHY IS $\mathcal{S}^{\rm cut}_{1/2}$ IS SMALL?

In this section, we discuss the well-established result that $\mathcal{S}^{\rm cut}_{1/2}$ is unexpectedly low (and much smaller than the full sky value), and consider the origin of this observation from a harmonic space perspective. Known aspects of the full sky realization are found to be behind the result, namely (a) the low amplitude, planarity and rough Galactic alignment of the quadrupole; and (b) the planarity and alignment of the octupole.

Combining equations (6) and (9) shows that the value of $\mathcal{S}^{\rm cut}_{1/2}$ derived from pixel-pair averages on the cut sky is:

$$\mathcal{S}^{\rm cut}_{1/2} = \sum_{\ell\ell'} \hat{C}^{\rm PCL}_\ell \hat{C}^{\rm PCL}_{\ell'} s_{\ell\ell'}. \qquad (10)$$

It is not clear why various groups have seen differences (albeit minor) in $\mathcal{S}^{\rm cut}_{1/2}$ measured numerically on the sky and $\mathcal{S}^{\rm cut}_{1/2}$ defined by equation (10), but it probably relates to difficulties in designing stable numerical schemes which estimate $\mathcal{C}(\theta)$ directly from pixel pairs.

To understand the origin of the observed $\mathcal{S}^{\rm cut}_{1/2}$ value, let us first note that, in a typical realization of the isotropic ΛCDM model, the primary driver of the ratio $\mathcal{S}^{\rm cut}_{1/2}/\mathcal{S}^{\rm full}_{1/2}$ is the ratio $\hat{C}^{\rm PCL}_2/\mathcal{C}^{\rm full}_2$. This reflects the fact that, substituting the theory C_ℓ's in equation (6), the dominant contribution is from the term quadratic in C_2. However, due to the low amplitude of the full sky quadrupole in our particular realization, the $\ell = 2$ mode becomes subdominant in determining $\mathcal{S}_{1/2}$. C_3 then gives the dominant

[2] This interpretation of the QML estimator's operation is a harmonic-space equivalent to the pixel-space 'high-pass filter' interpretation given by Tegmark [19].

[3] Closely related to this is the better-known result that PCL estimators for high ℓ's are nearly optimal for moderate sky cuts. Such cuts imply that the reconstruction is only sensitive to a finite window $\ell \pm \Delta\ell$, over which the high-ℓ ΛCDM power spectrum is nearly flat.

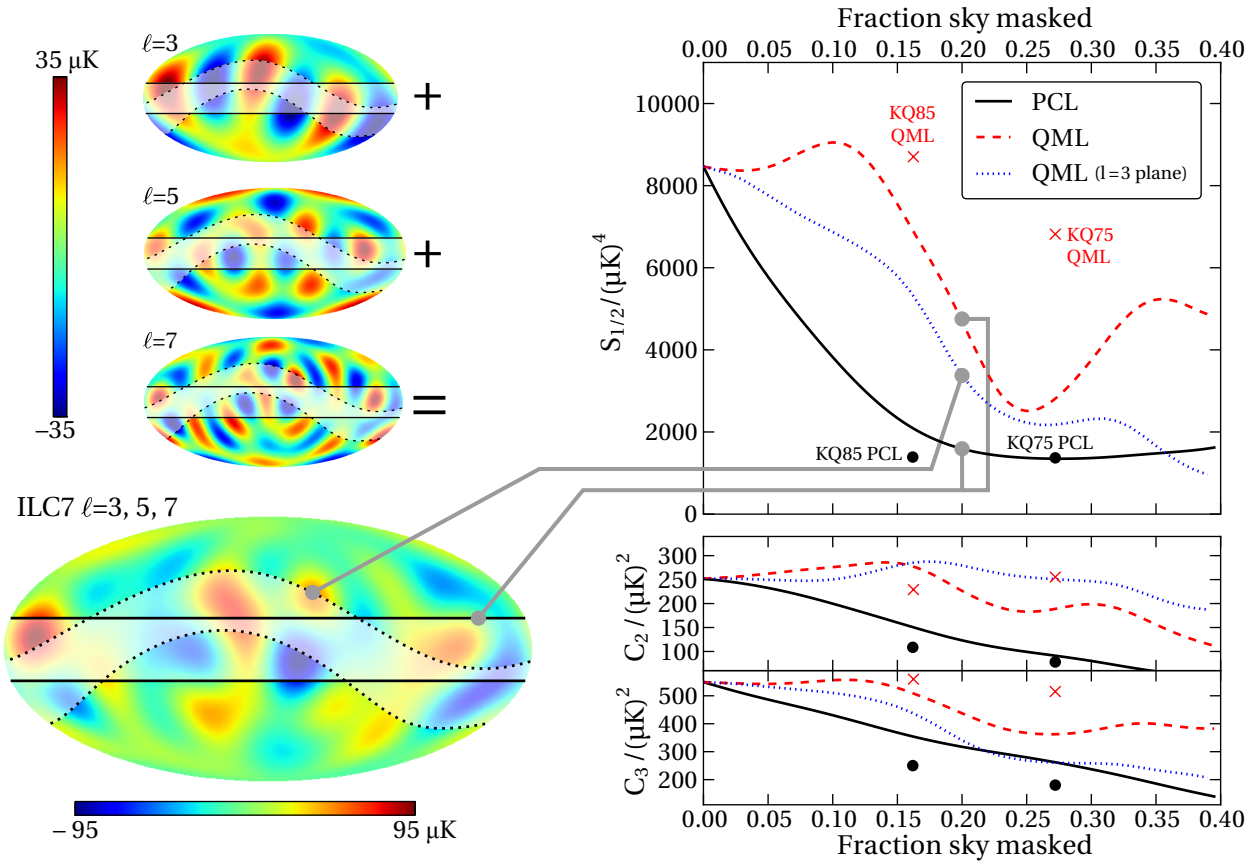

FIG. 1: (Color online.) An illustration of the low value of $\mathcal{S}_{1/2}^{\rm cut}$ and its origin. The upper left-hand panels show the pattern of ILC7 temperature fluctuations in $\ell = 3, 5$ and 7 modes from top downwards; the larger panel underneath shows the sum, in which anticorrelations between the modes cause power near the poles to be small. The shaded regions bounded by solid and dashed lines represent a 20% azimuthal mask in the Galactic and $\ell = 3$ angular momentum planes respectively (defined in text). The large uppermost panel on the right hand side shows the values of $\mathcal{S}_{1/2}^{\rm cut}$ for a variable width Galactic azimuthal mask (solid line). The dashed line shows the same result using QML, rather than PCL, reconstruction techniques; the dotted line shows the QML result when the mask is applied in the $\ell = 3$ angular momentum plane, demonstrating that full-sky power can be efficiently hidden even from the QML estimator. The two right-hand panels underneath the main plot show the corresponding estimated values $\hat{C}_2$ and $\hat{C}_3$. Plotted points show the results from using the WMAP team's KQ85y7 and KQ75y7 masks for PCL (circles) and QML (crosses) estimators respectively, exhibiting the QML estimator's relative sensitivity to mask shape.

contribution on the full sky, so that $\hat{C}_3^{\rm PCL}/\mathcal{C}_3^{\rm full}$ determines the magnitude of $\mathcal{S}_{1/2}^{\rm cut}/\mathcal{S}_{1/2}^{\rm full}$.

Now the octupole of the observed realization happens to be somewhat planar (although, as quantified below, not 'anomalously' so). Its preferred plane is, in turn, very roughly aligned with the Galactic plane. When the octupole is masked using one of the standard WMAP temperature analysis masks such as KQ75y7 (usable sky fraction $f_{\rm sky} = 70.6\%$) or KQ85y7 ($f_{\rm sky} = 78.3\%$), a significant amount of power in the octupole is hidden due to this approximate alignment, leading the recovered $\hat{C}_3^{\rm PCL}$ to be an under-estimate of the full sky value. Thus, according to the considerations above, $\mathcal{S}_{1/2}^{\rm cut}$ drops sharply in response.

Figure 1 illustrates this further by showing (top plot, solid line) the value of $\mathcal{S}_{1/2}^{\rm cut}$ derived from the seventh-year WMAP [22] Internal Linear Combination map (ILC7) as the sky coverage of an equatorial, azimuthal mask is increased from 0 to 40% (*i.e.* $f_{\rm sky}$ drops from 100% to 60%). When the sky is unmasked, the PCL and QML power spectrum reconstructions reduce to the full-sky estimate, so that all results agree for the nil cut. For the moment we will focus on the behaviour of the PCL reconstructions, returning to the QML cases (dashed and dash-dotted lines) momentarily.

As described above, the rapid decline in $S_{1/2}^{\rm cut}$ as a function of increasing mask width is largely due to the corresponding decline in $\hat{C}_3^{\rm PCL}$ (illustrated in the lowermost right-hand panel of Figure 1), which in turn can be linked to the progressive masking of the planar-concentrated power (see also the 20% sky-cuts illustrated in the CMB projections on the left of the Figure). The known planarity of the quadrupole is also important, in that the cut-sky estimates $\hat{C}_2^{\rm PCL}$ decline with increasing mask

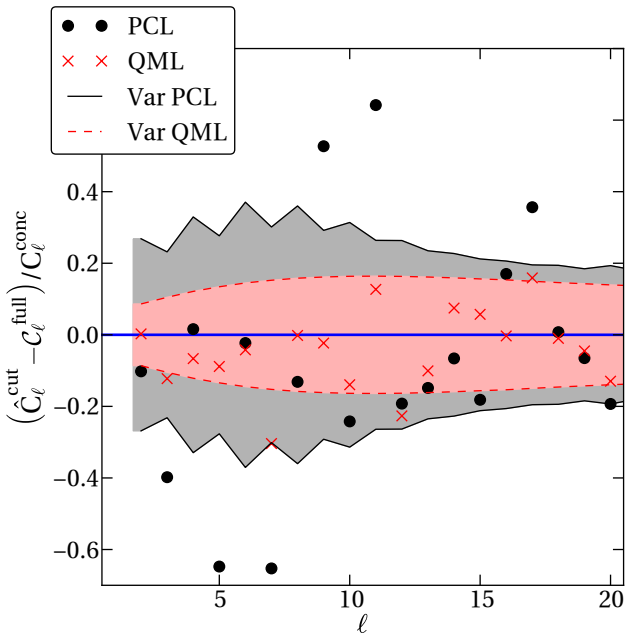

FIG. 2: (Color online.) The difference between the cut-sky estimates $\hat{C}_\ell$ and the full-sky measured values $\mathcal{C}_\ell$ on the ILC7 map, expressed for ease of viewing as a fraction of the fiducial theoretical best-fit C_ℓ's published by the WMAP team [21]. The results from a PCL and QML estimator for a 20% azimuthal sky-cut are shown by dots and crosses respectively. The shaded bands show the expected deviation of the cut-sky from the full-sky values. The larger, outer band represents the PCL deviation while the smaller, inner band (with dotted edges) represents the QML deviation. The low value of $\mathcal{S}_{1/2}^{\rm cut}$ arises because the PCL power spectrum reconstruction for $l \leq 8$ typically falls below, never significantly above, the full sky value.

area (central right-hand panel) and so do not regain dominance over the octupole contribution.

The described properties of multipoles $\ell \leq 3$ are not quite enough, on their own, to account for the low $\mathcal{S}_{1/2}^{\rm cut}$. Working with an azimuthal mask of 20%, if we use the cut sky values $\hat{C}_\ell^{\rm PCL}$ for $\ell \leq 3$ and the full sky values $\mathcal{C}_\ell$ for $\ell > 3$, we calculate a value of $\mathcal{S}_{1/2}^{\rm hybrid} = 3327\ \mu{\rm K}^4$. The true cut-sky value for the same mask is $\mathcal{S}_{1/2}^{\rm cut} = 1529\ \mu{\rm K}^4$. We can account for this discrepancy by noting that the PCL reconstructions of two other multipoles, $\ell = 5$ and $\ell = 7$, are also rather low.

The overall situation is illustrated in Figure 2, where we have plotted (as circles) the difference between full-sky and cut-sky power spectra $(\hat{C}_\ell^{\rm PCL} - \mathcal{C}_\ell)$ at $f_{\rm sky} = 80\%$; these are calculated from the ILC7 map and scaled by the concordance theory C_ℓ's [21]. The differences at $\ell = 3, 5, 7$ are all somewhat outside the 1σ variance (illustrated by the grey jagged band)[4].

The reason for the shortfall in reconstructed power in $\ell = 5$ and 7 is not immediately clear from inspecting their individual patterns on the sky (small panels near top left of Figure 1). Only when all the odd multipoles at $\ell \leq 7$ are summed does the power become visually planar (see the larger Mollweide projection at the bottom left of Figure 1). Thus cancellations between the $\ell = 3, 5, 7$ modes in the polar regions effectively hide power from estimators once the sky is masked. (We note that even-ℓ modes have no effect on the odd-ℓ reconstruction and vice versa, since these are decoupled when adopting an equatorially symmetric mask.)

A. Behaviour of the QML estimator

Having established the origin of the low $\mathcal{S}_{1/2}^{\rm cut}$, let us turn to the effect of using a QML, rather than PCL, estimator in reconstructing the full sky. It has been commented elsewhere [16] that, for KQ85y7 masks, the QML estimator reconstructs most of the power in the full sky octupole; $\mathcal{S}_{1/2}^{\rm QML}$ is close to $\mathcal{S}_{1/2}^{\rm full}$ even for the larger KQ75y7 mask. We have reproduced these results; QML outputs are plotted as crosses in the panels of Figures 1 and 2. As discussed above (Section II), the QML estimator re-weights its input maps to extract full-sky information as efficiently as possible; hence the improvement is not surprising. Recall that, even when considering anisotropic theories, the QML estimates for the full sky are expected to be superior (this is further reinforced in Section IV below).

However, the QML estimator's ability to make an efficient recovery of our full sky results does depend on the shape of the mask[5]. For an azimuthal mask covering a sky fraction $\gtrsim 20\%$, even the QML estimator starts to underestimate the power on the full sky (illustrated by the dotted lines in the right panels of Figure 1; the lower panel shows that the falling $\mathcal{S}_{1/2}^{\rm QML}$ tracks a drop in the power of the reconstructed octupole $\hat{C}_3^{\rm QML}$). These results show that, if the power is sufficiently localized within the mask, it cannot be reconstructed by any technique.

This interpretation of the results is confirmed by applying an azimuthal mask in the plane (as defined below) of the octupole (dotted lines in all panels of Figure 1).

---

[4] The standard deviation illustrated in Figure 2 is defined as $\langle(\hat{C}_\ell - \mathcal{C}_\ell)^2\rangle^{1/2}$, *i.e.* it is the 'cut-induced' variance introduced in Appendix A 3, equation (A15). We should note in passing that the variance is close to diagonal – *i.e.* correlations between the estimates for different ℓ are small in both PCL and QML cases.

[5] The PCL estimator is less sensitive to the exact shape of the cut than the QML estimator; this is to be expected given the simplicity of the former method (which, to a close approximation, measures the power on the cut sky and scales it by the appropriate sky fraction).

The octupole plane is defined by rotating the map until the "angular momentum dispersion" statistic [23] for quantifying the planarity of multipole ℓ is maximized:

$$L_\ell^2 = \frac{\sum_{m=-\ell}^{\ell} m^2 |a_{\ell m}|^2}{\ell^2 \sum_{m=-\ell}^{\ell} |a_{\ell m}|^2} \ . \tag{11}$$

(This maximization is achieved using a fast method described in Appendix E.) Now the octupole is masked from the map very efficiently, and $\hat{C}_3^{\rm QML}$ drops sharply as a consequence. In response, $\mathcal{S}_{1/2}^{\rm QML}$ becomes a severe underestimate, at large sky cuts becoming even worse than the Galactic azimuthal-masked PCL estimator.

B. Summary of the frequentist result

The preceding material has shown that the small measured value of $\mathcal{S}_{1/2}^{\rm cut}$ is attributable to a series of somewhat unlikely aspects of the observed realization. We now recap and discuss briefly the frequentist statistical significance of $\mathcal{S}_{1/2}^{\rm cut}$.

The primary contribution is the low $(254\,\mu{\rm K}^2)$ quadrupole amplitude (with a P-value of 4% given the best-fit power spectrum). The planarity of the octupole can be assessed by considering the rotation-maximized value of Eq. (11), which on the ILC7 map is 0.926. The P-value computed from 10,000 isotropic realizations for observing $L_3^2 > 0.926$ is $\sim 15\%$, *i.e.* our realization is not particularly unusual. The approximate alignment of this somewhat planar octupole with the Galactic cuts typically used in CMB data analyses can reasonably be regarded as purely coincidental (P-value 21%). A consistent picture is found from assessing Figure 2, in which the $\ell = 3$ PCL reconstruction deviates from the mean by about -1.5σ. Similarly the shortfall of PCL-reconstructed power in the $\ell = 5$ and 7 modes is a fluctuation of around -2σ. None of these observations on their own look particularly unusual; the statistical anomaly arises instead because *all* of the low-ℓ PCL estimates are low. Despite its suboptimal nature the PCL estimator is unbiased, and the reconstructions at different ℓ's are only weakly correlated, so one would have expected as many over- as under-estimates.

The frequentist significance of the $\mathcal{S}_{1/2}^{\rm cut}$ result is connected, then, to a series of coincident minor anomalies in our realization. Only when combined in a specific way do these observations raise frequentist alarm. Of course, this simply shows that we have found a way to 'factor' the low P-value of $\mathcal{S}_{1/2}^{\rm cut}$, which does not by itself determine whether the anomaly might point to theories beyond the concordance model. Therefore in the next Section, we consider the feasibility of finding theories which are statistically preferred to the concordance theory in a Bayesian comparison of $\mathcal{S}_{1/2}^{\rm cut}$.

IV. ANISOTROPIC THEORIES

We have shown in the previous Section that the low observed $S_{1/2}^{\rm cut}$ can be attributed to the fortuitous alignment of power in the $\ell = 3, 5$ and 7 modes of the CMB (along with the planarity, and small full-sky amplitude, of the quadrupole).

Broadly, one can imagine three distinct ways in which the small observed cut-sky power in the $\ell = 3, 5$ and 7 modes could look less anomalous in an alternative theory:

1. The $\hat{C}_\ell^{\rm PCL}$ estimates could turn out to be biased in the ensemble mean of the true theory;
2. The $\hat{C}_\ell^{\rm PCL}$ estimates could have a larger variance in the ensemble of the true theory, making the departures from the mean less significant;
3. The true theory could correlate $\hat{C}_\ell^{\rm PCL}$ estimates so that the likelihood of small cut-sky power in $\ell = 5, 7$ is greater once the small cut-sky power in $\ell = 3$ is known.

Our main focus in what follows will be on (1); possibilities (2) and (3) will be mentioned where relevant.

Given a true covariance matrix $\mathbf{C}$ related to the concordance isotropic theory by $\mathbf{C} = \mathbf{C}^{\rm conc} + \mathbf{A}$, one may explicitly calculate the bias,

$$\text{Bias}_\ell = \langle \hat{C}_\ell - \mathcal{C}_\ell \rangle \tag{12}$$

and the variance

$$V_{\ell\ell'} = \left\langle \left(\hat{C}_\ell - \langle \hat{C}_\ell \rangle \right) \left(\hat{C}_{\ell'} - \langle \hat{C}_{\ell'} \rangle \right) \right\rangle , \tag{13}$$

where the false covariance matrix $\mathbf{C}^{\rm conc}$ is used in constructing estimators, but the true covariance matrix $\mathbf{C}$ is employed in taking the final ensemble average. For algebraic expressions the reader is referred to Appendix A 3.

To gain a feel for how alternative theories can influence the recovered power spectrum on the cut sky, let us consider the following specific cases.

(i) *Galactic contamination*, *i.e.* residual errors in the Galactic signal subtraction. This is modelled by creating a template map of possible errors, taking 1% of the difference of the WMAP7 ILC map with the V-band map (after smoothing to a common resolution of 1°). The map gives us a rough handle on the form of the residual contamination to be expected (albeit with an unknown amplitude). In the ensemble, the template map is simply added to the observed CMB sky, yielding equivalent results to a theory with anisotropic Gaussian correction $\mathbf{A} = \mathbf{g}\mathbf{g}^\dagger$, where $\mathbf{g}$ represents the spherical harmonic coefficients of the contamination map. It has already been shown in Ref. [24] that this kind of contamination cannot improve the likelihood of $\mathcal{S}_{1/2}^{\rm cut}$, but the model remains helpful for our discussion below.

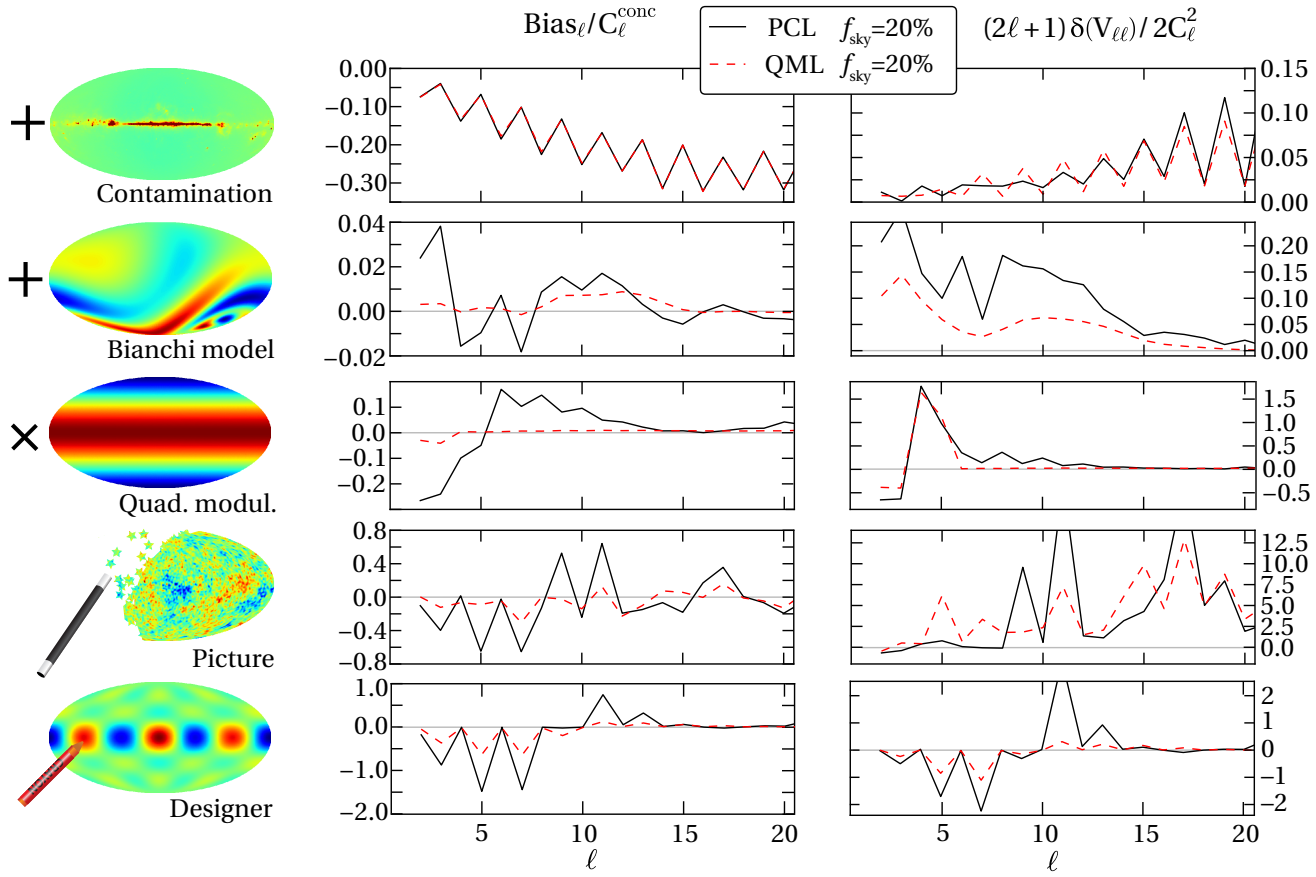

FIG. 3: (Color online.) For various anisotropic theories we plot the biases (left panels) and change in the diagonal part of the variance (right panels) compared to the isotropic case. In this context non-zero 'bias' can be desirable, for instance in the case illustrated in the uppermost panels, where it merely reflects that the full sky power spectrum is contaminated by the presence of the galaxy. The solid lines show statistics for PCL reconstructions of a $f_{\rm sky} = 20\%$ azimuthally masked sky; the dashed lines show the corresponding QML reconstructions. As expected, the QML reconstructions are generally more accurate (despite the anisotropy of the underlying theory); for more details see text. We scale the biases by the isotropic theory power spectrum and the change in the variances by the isotropic cosmic variance. The first three theories are simple existing models while the latter two are tests which are designed specially to reproduce the $\mathcal{S}_{1/2}^{\rm cut}$ anomaly. Note the different scales on each panel.

(ii) *Bianchi VII$_h$ template.* Using the algorithms of Refs. [25, 26] we calculate a temperature anisotropy template for the Bianchi VII$_h$ vector mode case with an amplitude of $35\,\mu$K according to the best-fit parameters of Ref. [27]. In the ensemble this is added to the concordance CMB as with the templates considered above. Physically, such a setup can be motivated by the existence of anisotropic Bianchi modes which are well behaved at the initial singularity, although such models are fine-tuned.

(iii) *Quadrupolar Modulation.* A strong quadrupolar modulation[6] of the temperature field is known to reproduce the co-planarity of quadrupole and octupole [30]. The modulation is required to have a very large amplitude, yet be confined to low multipoles. Dvorkin *et al.* [31] discuss how any early universe model of such a modulation must be carefully tuned in harmonic space to avoid the leakage of modulated power to high multipoles through projection effects. We approximate these considerations by modulating only the quadrupole and octupole of isotropic realizations. Since a quadrupolar modulation of a multipole ℓ couples power to $\ell \pm 2$, on the full sky our modulation only has an effect on multipoles $\ell \leq 5$.

[6] Attention has also been given in the past to dipolar modulations. In the case of an equatorial azimuthal mask this can have only second order effects on power spectrum reconstruction, since it couples ℓ to $\ell \pm 1$ while the mask couples ℓ to $\ell \pm 2n$. A different quadrupolar anisotropy, that of the inferred primordial power spectrum, has also been reported [12, 13, 28]. However, Hanson et al. [29] have identified WMAP beam asymmetries as the origin of this unconnected effect.

(iv) *Picture* and *Designer* theories. These are specific theories designed to investigate the best possible statistical gains to be made from anisotropic theories over the concordance case. We will describe them in detail in Section IV B, below.

The left panel of Figure 3 exhibits the biases induced by each of these theories [defined by equation (12)]; the right hand panel shows the diagonal part of the variance [defined by equation (13)]. In both panels, the results from the PCL estimators are plotted as solid lines, while the QML results are shown by dashed lines. The sky cut imposed for these calculations is a 20% Galactic azimuthal mask.

In the first case, that of Galactic contamination, the 'bias' reflects the added power from the Galaxy, visible in the full sky $\mathcal{C}_\ell$'s but naturally invisible to reconstructions made from the cut sky (in which the Galaxy is masked away). Both the PCL and QML estimators therefore become equally 'biased', but this is, in fact, a desirable feature: they are rejecting the contamination. Note that the zig-zag pattern in the bias arises from the rough equatorial symmetry of the Galaxy, which results in a much stronger coupling to even, rather than odd, ℓ's. On the other hand the zig-zag in the biases actually observed (Figure 2) is larger at odd, rather than even, ℓ's. Finally, the apparent biases and extra variance of spatially localized contamination tends to grow towards high ℓ as a fraction of the full sky power, whereas the observed discrepancies are confined to low ℓ.

The second case (Bianchi contamination) is similar in that it adds a template to the concordance covariance; but because the power is not localized within the mask, it is now visible even on the cut-sky. As expected from our earlier considerations, the QML estimator in this regime reconstructs the full sky $\mathcal{C}_\ell$ power with a smaller bias and variance than the PCL case. The increased variance (of order 20% of the cosmic variance) is more significant than the bias (of order 2% of the power spectrum). This can be understood by noting that, since the Bianchi signal has a small rms power of $\sim 12\,\mu\mathrm{K}^2$, the individual elements of the template covariance contribution $\mathbf{A}$ are much smaller than the elements of the concordance covariance matrix $\mathbf{C}^{\mathrm{conc}}$. Expanding an expression for the variance (A14) highlights the existence of cross-terms in $\mathbf{C}^{\mathrm{conc}}$ and $\mathbf{A}$; it is these leading order contributions which give the larger variance.

Let us now turn away from additive modifications to the concordance theory, and instead discuss the quadrupolar modulation. When analysed on the cut sky, the power at low multipoles is hidden from the PCL estimator (but less so from the QML estimator), leading to a negative bias. The modulation also couples $\ell \pm 2$, creating power on the full sky in $\ell = 4, 5$; this accounts for the spike at these multipoles in the variance of the estimators. The extra power is further spread to higher ℓ by mode-coupling resulting from the masking operation. The result is that the PCL estimator over-estimates power in multipoles $5 < \ell < 15$; note that, because the power spectrum is rapidly decreasing, a small leakage of power to high ℓ from the low multipoles can result in a substantial bias. Once again the QML estimator fares better, more efficiently confining the contamination to low ℓ.

The quadrupolar modulation behaves qualitatively as expected, hiding power at low ℓ; this is the right sort of effect to reproduce the low $\mathcal{S}_{1/2}^{\mathrm{cut}}$ and therefore produces a small increase in the likelihood of the observed value. However, Figure 3 shows that the biases from this theory are rather small. Therefore, rather than focus on this model, we can go one stage further and consider tuning Gaussian models to fit the value of $\mathcal{S}_{1/2}^{\mathrm{cut}}$ as closely as possible. These theories are less transparent in their physical meaning, but are guaranteed to give a better fit to the observed properties of the sky.

A. Picture theory

Consider the theory which exactly matches the observed CMB; it has covariance matrix $\mathbf{C} = \mathbf{a}\mathbf{a}^\dagger$ where $\mathbf{a}$ represents the observed ILC $a_{\ell m}$'s. Because $\mathbf{C}$ has zero variance in any direction orthogonal to the observed data, it has an infinite likelihood (or, more correctly, a likelihood bounded from above only by noise in the experiment).

The ensemble for this theory is a series of pictures of our own CMB sky (represented here by the ILC map), scaled by a Gaussian random amplitude of unit variance. Consequently the biases exactly match the values of $\hat{C}_\ell - \mathcal{C}_\ell$ for our observed sky (see Figure 3). However, because there is only one mode (the amplitude scaling of the entire sky), the variances become extremely large. To build intuition, consider the cosmic variance of the concordance model, in which the variance on $\mathcal{C}_\ell$ decreases as $2\ell+1$. This arises solely because of the additional modes available at increasing ℓ; in the picture theory all modes are perfectly correlated, so the cosmic variance does not decline in this way.

In spite of the divergently large likelihood for the picture theory, the variance means that our observed value of $\mathcal{S}_{1/2}^{\mathrm{cut}}$ has a finite likelihood which can be calculated by Monte Carlo simulation of the ensemble. In Figure 4 we plot the log likelihood for the concordance ΛCDM model (solid line), the picture theory (dash-dotted line) and the designer theory (dashed line), the last of which we will return to momentarily.

The improvement in the log likelihood of $\mathcal{S}_{1/2}^{\mathrm{cut}}$ for the picture model (over the isotropic concordance case) is $\Delta \ln \mathcal{L} = 3.7$. This disappointingly modest improvement can be seen to result from the large cosmic variance in a theory with only one degree of freedom: while it peaks near the observed value, the $\mathcal{S}_{1/2}^{\mathrm{cut}}$ likelihood function for the picture theory is extremely broad. The broadness in turn impacts upon the peak value because the likelihood must be normalized to one.

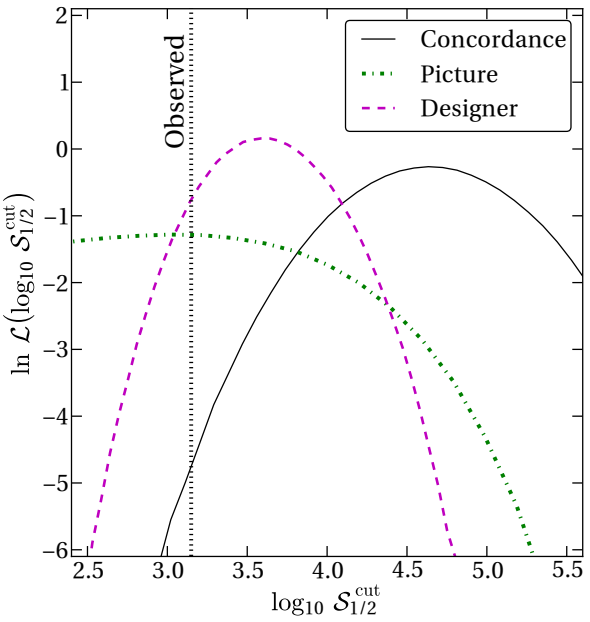

FIG. 4: (Color online.) The likelihoods for $\mathcal{S}_{1/2}^{\rm cut}$ compared between three theories: the isotropic concordance theory (solid line), the 'designer' theory (dashed line) and 'picture' (dash-dotted line) theory. The latter two are specifically designed to reproduce low $\mathcal{S}_{1/2}^{\rm cut}$. The improvement in the log likelihood over the concordance cases are respectively 4.2 and 3.7, which are very small improvements given the fine-tuning involved. Since these theories should produce the greatest possible gains in likelihood, the value of the observed $\mathcal{S}_{1/2}^{\rm cut}$ statistic is not a strong objection to the concordance theory.

B. Designer theory

We have examined the picture theory in which the likelihood of our own observed CMB is, by design, divergently large. The $\mathcal{S}_{1/2}^{\rm cut}$ likelihood was shown to barely favour the theory because the variance on this value is so large. In this section we search for a theory with similar properties to the observed sky, but allow power to be spread through many more modes, so that the variance in high-order statistics such as $\mathcal{S}_{1/2}^{\rm cut}$ is better controlled.

A full explanation of the mathematical construction is given in Appendix C. Our search is over all positive definite covariance matrices $\mathbf{C}$, corresponding to all Gaussian theories with zero mean ($\langle \Delta T \rangle = 0$), subject to two sets of constraints. Firstly, the covariance matrix is required to have full-sky theory C_ℓ's [defined by Eq. (3)] equal to the observed ILC values ($C_\ell = \mathcal{C}_\ell^{\rm ILC}$). Secondly, the PCL estimator applied to the theory on the 20% azimuthally masked sky is required to find zero power in $\ell = 2, 3, 5$ and 7 (*i.e.* Bias$_\ell = -\mathcal{C}_\ell$ at these ℓ's). In order to satisfy these results simultaneously, our technique naturally introduces anisotropic correlations between different multipoles. Substantial freedom remains, which we use to minimize the cosmic variance of the final theory (see Appendix C). The freedom is truncated at $\ell_{\rm max} = 10$; we adopted the concordance covariance for all $\ell > \ell_{\rm max}$, but verified that our conclusions are insensitive to this choice.

The CMB projection labelled 'Designer' in Figure 3 illustrates an actual realization from this model (although only the $\ell = 3, 5$ and 7 modes are plotted). One can see that the theory is very efficient at localizing power in modes confined within our specified 20% mask (while keeping the full sky power spectrum to the specified values). The bias panel for this theory shows that, accordingly, no power is detected by the cut-sky PCL estimator in $\ell = 3, 5$ and 7. (The plot shows that some of the biases are actually smaller than $-C_\ell$, which is as expected since $\mathcal{C}_\ell^{\rm ILC} > C_\ell$ at the corresponding multipoles.) At low ℓ, the variance on the $\hat{C}_\ell$'s is smaller than for the concordance model ($\delta V_{\ell\ell} < 0$), because the reconstructed power in these modes remains close to zero in all realizations. For $\ell \gtrsim 10$, there is a spike of larger variance arising from the mask-induced contamination similar to that described for the modulation model. At large ℓ, the variance tends to the standard concordance variance for the estimators ($\delta V_{\ell\ell} = 0$). Once again, the QML estimator performs better in minimizing both bias and variance compared to the PCL case at nearly all ℓ.

We can now return to Figure 4 which displays, as a dashed curve, the likelihoods for our designer theory. We described above how the power localization at low ℓ favours a low $\mathcal{S}_{1/2}^{\rm cut}$; accordingly the peak likelihood (at $\log_{10} \mathcal{S}_{1/2}^{\rm cut}/\mu{\rm K}^4 = 3.6$) is considerably smaller than the equivalent value for the concordance theory ($\log_{10} \mathcal{S}_{1/2}^{\rm cut}/\mu{\rm K}^4 = 4.6$). However, despite being minimized by spreading power through more degrees of freedom, the variance of the designer $\mathcal{S}_{1/2}^{\rm cut}$ remains large and consequently the improvement in likelihood is modest ($\Delta \ln \mathcal{L} = 4.2$) despite the dramatic increase in the number of degrees of freedom needed to construct this theory.

It is clear from Figure 4 that to obtain significant gains in likelihood for $\mathcal{S}_{1/2}^{\rm cut}$, one needs to achieve far smaller cosmic variance on $\mathcal{S}_{1/2}^{\rm cut}$. But the designer theory plausibly gives near the smallest possible variance on this quantity. In particular, Appendix C derives a minimum bound on the variance of $\mathcal{S}_{1/2}^{\rm cut}$. The lower bound can be understood as arising from a suitable isotropic limit (which is unattainable in practice, but provides a provable lower limit for attainable theories); isotropic theories minimize the cosmic variance for a given power spectrum, because they maximize the number of independent modes and spread power through these modes as evenly as possible. The lower bound calculated from the appropriate isotropic test-case is $\sigma^2_{\rm min}(\mathcal{S}_{1/2}^{\rm cut}) = 1.7 \times 10^7 \mu{\rm K}^8$, compared with the variance on the designer theory of $\sigma^2(\mathcal{S}_{1/2}^{\rm cut}) = 2.9 \times 10^7 \mu{\rm K}^8$. Thus the designer theory detailed in this section almost saturates the variance limit; we may be confident that no Gaussian theory can have a significantly more peaked likelihood.

C. Summary and discussion

The two theories we have discussed in the preceding two sections (the first giving an infinite likelihood for the observed sky; the second tuned as far as we can to give a large likelihood for the single value $\mathcal{S}^{\rm cut}_{1/2}$) strongly suggest that no anisotropic Gaussian theory can improve the likelihood of the observed $\mathcal{S}^{\rm cut}_{1/2}$ by more than $\Delta \ln \mathcal{L} \simeq 5$. Because of the careful fine-tuning of these models, they form a plausible upper bound for the statistical gain available.

How can we interpret this very modest likelihood gain? From a Bayesian perspective, correlations between the primary temperature (from the high-z last scattering surface) and the integrated Sachs-Wolfe signal (from local structure) are implied in any model aligning low-ℓ power [16, 31, 32]. Therefore any realistic physical prior probability is very small relative to the isotropic ΛCDM cosmogony; the posterior probability ratios will still vastly favour the latter theory.

An alternative argument is as follows. Up to $\ell_{\rm max} = 10$ (and excluding monopole and dipole), the anisotropic theory has approximately 6 900 degrees of freedom compared to the isotropic case with 8 degrees of freedom. Thus the improvement in log likelihood per degree of freedom is of order 10^{-3}. While this is not a strictly Bayesian interpretation, it does suggest that our statistical gain has been achieved only with enormous fine-tuning.

Both of these lines of reasoning suggest strongly that the observed value of $\mathcal{S}^{\rm cut}_{1/2}$ can never constitute strong evidence in favour of any Gaussian theory posited as an alternative to ΛCDM. One escape route from this result is to consider, rather than the absolute value of $\mathcal{S}^{\rm cut}_{1/2}$, the ratio $\mathcal{S}^{\rm cut}_{1/2}/\mathcal{S}_{1/2}$. In the case of the picture theory, for instance, one then has an infinite likelihood for the observed value of the ratio (because the one available random degree of freedom – the amplitude – cancels between numerator and denominator). This immediately demonstrates that there is no upper bound to the likelihood gain for such a statistic. We would suggest, however, that the existence of an upper bound for the original statistic, $\mathcal{S}^{\rm cut}_{1/2}$, is an attractive property – precisely because it allows for an understanding of the Bayesian theoretical improvements available without detailed physical modelling. The most convincing way to show that the observed sky is anomalous would therefore be to find a statistic encapsulating the planarity and correlation of power which has large, but not trivially infinite, likelihood gains available. Starting from the results of Section III, such a statistic might be developed from physical considerations on the full sky.

V. CONCLUSIONS

There is a classic difficulty in understanding large and complex datasets such as those produced by WMAP and, in the future, *Planck*: they contain so much information that statistical anomalies can be found without any difficulty. We have taken as an example the purported anomalous aspects of the angular correlation function. Some previous work claims that, after considering these anomalies, the entire cosmological paradigm is to be doubted [10]; other authors claim that apparent anomalies can be dismissed as the product of *a posteriori* analysis [16]. Yet *a posteriori* reasoning must be allowed in science, since otherwise we would rarely, if ever, recognize failings of our existing knowledge.

The contrary statistical claims relating to $S^{\rm cut}_{1/2}$ are reconciled by appreciating that, without an alternative theory to test against, there is no unambiguous significance to any anomaly. We have therefore presented an alternative approach to this puzzle: we examined the origin of the low $S^{\rm cut}_{1/2}$ in harmonic space, and then attempted to find theories that reproduce the required patterns.

In the process we noted that the cut-sky correlation function contains identical information to the PCL power spectrum estimates. We therefore used the PCL estimates for the majority of our results, but also demonstrated that the standard QML techniques provide more reliable reconstructions of the full sky, even when anisotropy is suspected. We informed our intuition about the behaviour of the estimators by considering simple anisotropic modifications to the concordance models (contamination, Bianchi and quadrupolar modulation theories). This showed explicitly that the QML estimator biases introduced by anisotropic theories were smaller than or comparable to the PCL case.

Then, by attempting to construct anisotropic Gaussian theories which improve the likelihood of the low $\mathcal{S}^{\rm cut}_{1/2}$, we demonstrated that no significant gains in likelihood for this single statistic are available. Since there is no suggestion in the observed sky that the underlying ensemble is significantly non-Gaussian [33], it is implausible that post-Gaussian corrections would substantially change our results. We therefore conclude that the $\mathcal{S}^{\rm cut}_{1/2}$ anomaly is not likely to point to new physics.

If it does have any meaning, the $\mathcal{S}^{\rm cut}_{1/2}$ anomaly (and the underlying shortfall of power seen by PCL estimators) does not indicate a vanishing large-scale correlation function, but rather is related to alignments of low-ℓ power on the full sky (Section III). It is likely that full-sky statistics can be constructed which capture these unexpected correlations better than $\mathcal{S}^{\rm cut}_{1/2}$ – and these could evade our likelihood limits. However, we argued that more trivial modifications (such as taking the ratio $\mathcal{S}^{\rm cut}_{1/2}/\mathcal{S}_{1/2}$) which sidestep our constraint by attaining an infinite likelihood under the 'picture' theory ($\mathbf{C} = \mathbf{a}\mathbf{a}^\dagger$) are not helpful; see Section IV C. In other words it is highly desirable to choose statistics, such as $\mathcal{S}^{\rm cut}_{1/2}$, that do allow for a finite limit to be placed on the Bayesian statistical gain available under a wide class of alternative straw-man models. Considering the magnitude of that limit is then, in our view, a plausible way to probe the significance of *a posteriori* anomalies.

Acknowledgments

We thank Anthony Challinor, Steven Gratton, Daniel Mortlock, Antony Lewis, Wayne Hu, George Efstathiou, Glenn Starkman, Dragan Huterer and Dominik J. Schwarz for productive discussions. AP is supported by Emmanuel College, Cambridge. HVP is supported by Marie Curie grant MIRG-CT-2007-203314 from the European Commission, and by STFC and the Leverhulme Trust. HVP thanks the Aspen Center for Physics for hospitality. We acknowledge use of the Healpix package [34] and the Legacy Archive for Microwave Background Data Analysis (LAMBDA). Support for LAMBDA is provided by the NASA Office of Space Science.

[1] C. L. Bennett *et al.*, ApJS **148**, 1 (2003).
[2] N. Jarosik *et al.*, ApJS (Submitted), 1001.4744.
[3] C. J. Copi, D. Huterer, and G. D. Starkman, Phys. Rev. **D70**, 043515 (2004), astro-ph/0310511.
[4] H. K. Eriksen, F. K. Hansen, A. J. Banday, K. M. Gorski, and P. B. Lilje, Astrophys. J. **605**, 14 (2004), astro-ph/0307507.
[5] A. de Oliveira-Costa, M. Tegmark, M. Zaldarriaga, and A. Hamilton, Phys. Rev. **D69**, 063516 (2004), astro-ph/0307282.
[6] K. Land and J. Magueijo, Physical Review Letters **95**, 071301 (2005), astro-ph/0502237.
[7] A. de Oliveira-Costa and M. Tegmark, Phys. Rev. D **74**, 023005 (2006), astro-ph/0603369.
[8] C. J. Copi, D. Huterer, D. J. Schwarz, and G. D. Starkman, Phys. Rev. D **75**, 023507 (2007), astro-ph/0605135.
[9] F. K. Hansen, A. J. Banday, K. M. Gorski, H. K. Eriksen, and P. B. Lilje, Astrophys. J. **704**, 1448 (2009), 0812.3795.
[10] C. J. Copi, D. Huterer, D. J. Schwarz, and G. D. Starkman, MNRAS **399**, 295 (2008), 0808.3767.
[11] C. L. Bennett *et al.*, ApJS (Submitted), 1001.4758.
[12] N. E. Groeneboom, L. Ackerman, I. Kathrine Wehus, and H. K. Eriksen, ApJ (Submitted), 0911.0150.
[13] N. E. Groeneboom and H. K. Eriksen, ApJ **690**, 1807 (2009), 0807.2242.
[14] H. Zheng and E. F. Bunn, Phys. Rev. D (Submitted), 1003.5548.
[15] E. T. Jaynes and G. L. Bretthorst, *Probability Theory* (Cambridge, UK: Cambridge University Press, 2003).
[16] G. Efstathiou, Y. Ma, and D. Hanson, MNRAS (Submitted), 0911.5399.
[17] N. Afshordi, Y. Loh, and M. A. Strauss, Phys. Rev. D **69**, 083524 (2004), astro-ph/0308260.
[18] B. D. Wandelt, E. Hivon, and K. M. Górski, Phys. Rev. D **64**, 083003 (2001), astro-ph/0008111.
[19] M. Tegmark, Phys. Rev. D **55**, 5895 (1997), astro-ph/9611174.
[20] G. Efstathiou, MNRAS **349**, 603 (2004), astro-ph/0307515.
[21] D. Larson *et al.*, ApJS (Submitted), 1001.4635.
[22] B. Gold *et al.*, ApJS (Submitted), 1001.4555.
[23] A. de Oliveira-Costa *et al.*, Phys. Rev. D **68**, 083003 (2003), astro-ph/0212419.
[24] E. F. Bunn and A. Bourdon, Phys. Rev. D **78**, 123509 (2008), 0808.0341.
[25] A. Pontzen and A. Challinor, MNRAS **380**, 1387 (2007), 0706.2075.
[26] A. Pontzen and A. Challinor, in preparation (2010).
[27] T. R. Jaffe, A. J. Banday, H. K. Eriksen, K. M. Górski, and F. K. Hansen, ApJ **643**, 616 (2006), astro-ph/0603844.
[28] D. Hanson and A. Lewis, Phys. Rev. D **80**, 063004 (2009), 0908.0963.
[29] D. Hanson, A. Lewis, and A. Challinor, Phys. Rev. D (Submitted), 1003.0198.
[30] C. Gordon, W. Hu, D. Huterer, and T. Crawford, Phys. Rev. D **72**, 103002 (2005), astro-ph/0509301.
[31] C. Dvorkin, H. V. Peiris, and W. Hu, Phys. Rev. D **77**, 063008 (2008), 0711.2321.
[32] C. L. Francis and J. A. Peacock, MNRAS in press (2010), 0909.2495.
[33] E. Komatsu *et al.*, ApJS (Submitted), 1001.4538.
[34] K. M. Górski *et al.*, ApJ **622**, 759 (2005), astro-ph/0409513.
[35] I. Szapudi, S. Prunet, D. Pogosyan, A. S. Szalay, and J. R. Bond, ApJ **548**, L115 (2001), astro-ph/0010256.
[36] G. Chon, A. Challinor, S. Prunet, E. Hivon, and I. Szapudi, MNRAS **350**, 914 (2004), astro-ph/0303414.
[37] D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskii, *Quantum theory of angular momentum* (Teaneck, NJ: World Scientific, 1987).

Appendix A: Quadratic Estimators: some useful results

In this Appendix, we summarize some technical details omitted from the main paper alongside useful results pertaining to the two most common quadratic power spectrum estimators: the pseudo-C_ℓ and quadratic maximum likelihood estimators. These will be introduced and compared in a single quadratic estimator framework to gain insights into their similarities and differences. Since we work in harmonic space, we first explain the sky-masking operation.

1. Cutting the sky

It is standard practice in CMB analysis to remove regions of the sky in which contamination from the Galaxy (or other undesirable sources) is suspected. This is accomplished by masking the temperature field and then constructing measurements based solely on the masked data. In harmonic space, the masked temperature expansion coefficients $\tilde{a}_{\ell m}$ are related to the unmasked $a_{\ell m}$

via

$$\tilde{a}_{\ell m} = K_{\ell m \ell' m'} a_{\ell' m'}, \text{ where} \quad \text{(A1)}$$

$$K_{\ell m \ell' m'} = \int \mathrm{d}\Omega\, Y^*_{\ell m}(\Omega) Y_{\ell' m'}(\Omega) M(\Omega). \quad \text{(A2)}$$

Here $M(\Omega)$ is 0 within the masked region and 1 outside. [See Appendix D for a brief discussion of a hidden numerical pitfall in equation (A2).] We can write expression (A1) compactly as the linear transformation

$$\tilde{\mathbf{a}} = \mathbf{K}\mathbf{a}, \quad \text{(A3)}$$

where $\mathbf{a}$ is a vector composed of the $a_{\ell m}$'s.

It will be helpful to note that $\mathbf{K}$ is both idempotent ($\mathbf{K}^2 = \mathbf{K}$) and Hermitian ($\mathbf{K}^\dagger = \mathbf{K}$). These identities may both be derived straight-forwardly from equation (A2); together they allow much flexibility in manipulating certain equations. In these appendices, the addition of a tilde will represent masked quantities and operators; thus $\tilde{\mathbf{b}} = \mathbf{K}\mathbf{b}$ for any data vector $\mathbf{b}$, while for any matrix $\mathbf{M}$ we write

$$\tilde{\mathbf{M}} = \mathbf{K}\mathbf{M}\mathbf{K}, \quad \text{(A4)}$$

implicitly taking advantage of the Hermitian property. For most of our numerical results we have assumed the mask is azimuthally symmetric, $M(\theta, \phi) = M(\phi)$. This is a reasonable approximation to true Galactic masks, and results in enormous computational simplification because $\mathbf{K}$ becomes sparse,

$$K_{\ell m \ell' m'} = K^m_{\ell \ell'} \delta_{m m'} \text{ (no sum)}. \quad \text{(A5)}$$

However all algebraic results are obtained with no such assumptions and are applicable to any type of mask.

Since the estimators considered here are quadratic in the cut-sky $\tilde{a}_{\ell m}$'s, we may write for a generic estimate $\hat{C}_\ell$:

$$\hat{C}_\ell = \tilde{\mathbf{a}}^\dagger \mathbf{R}^\ell \tilde{\mathbf{a}} \quad \text{(A6)}$$

for some set of matrices $\mathbf{R}^\ell$. Before explicitly defining these matrices, we describe a helpful notational trick and discuss a couple of generic features of quadratic estimators.

2. A helpful notational trick

Recall that in Sec. II the power spectrum $\mathcal{C}_\ell$ observed in our single realization of the full sky was defined as

$$\mathcal{C}_\ell \equiv \frac{1}{2\ell+1} \sum_m |a_{\ell m}|^2, \quad \text{(A7)}$$

and the theoretical power spectrum C_ℓ was taken to be the expectation value of Eq. (A7), $C_\ell = \langle \mathcal{C}_\ell \rangle$. We will henceforth use a shorthand for such expressions, writing

$$\mathcal{C}_\ell \equiv \frac{a^\dagger \mathbf{\Delta}^\ell a}{2\ell+1}; \quad \text{(A8)}$$

$$C_\ell \equiv \frac{\mathrm{Tr}\ \mathbf{C}\mathbf{\Delta}^\ell}{2\ell+1}, \quad \text{(A9)}$$

where $\mathbf{C} = \langle a a^\dagger \rangle$ is the theory covariance matrix and the elements of the $\mathbf{\Delta}^\ell$ matrices are

$$\left(\mathbf{\Delta}^\ell\right)_{\ell' m', \ell'' m''} = \delta^\ell_{\ell'} \delta^\ell_{\ell''} \delta_{m' m''} \text{ (no sum)}. \quad \text{(A10)}$$

Thus $\mathbf{\Delta}^\ell$ is the projection operator into the spin-ℓ subspace. The following two properties of $\mathbf{\Delta}^\ell$ are useful:

$$\mathrm{Tr}\ \mathbf{\Delta}^\ell = (2\ell+1), \quad \text{(A11)}$$

$$\mathbf{\Delta}^\ell \mathbf{\Delta}^{\ell'} = \delta_{\ell\ell'} \mathbf{\Delta}^\ell \text{ (no sum)}. \quad \text{(A12)}$$

Introducing the set of matrices $\mathbf{\Delta}^\ell$ produces considerably more compact and readable equations at later stages.

3. Expectation and variances

Given the cut sky power spectrum estimates $\hat{C}_\ell$ defined by equation (A6), we have respectively

$$\langle \hat{C}_\ell \rangle = \mathrm{Tr}\ \tilde{\mathbf{C}} \mathbf{R}^\ell \quad \text{(A13)}$$

$$V_{\ell\ell'} \equiv \langle \hat{C}_\ell \hat{C}_{\ell'} \rangle - \langle \hat{C}_\ell \rangle \langle \hat{C}_{\ell'} \rangle = 2\,\mathrm{Tr}\ \tilde{\mathbf{C}} \mathbf{R}^\ell \tilde{\mathbf{C}} \mathbf{R}^{\ell'} \quad \text{(A14)}$$

for the expectation and variance, where $\tilde{\mathbf{C}} = \langle \tilde{a} \tilde{a}^\dagger \rangle = \mathbf{K}\mathbf{C}\mathbf{K}$ is the cut-sky harmonic covariance matrix.

The estimator variance $V_{\ell\ell'}$ characterizes the random error associated with estimating the ensemble quantity C_ℓ from a single masked realization. This is the appropriate quantity for most results in the paper and appendices. However, occasionally one wants a measure of the extent to which the cut sky estimators accurately predict the full sky (rather than ensemble averaged) power. A suitable quantification is given by the following, which might be termed the 'cut-induced variance' (since it is necessarily zero on the full sky):

$$\mathrm{CIV}_{\ell\ell'} \equiv \left\langle \left(\hat{C}_\ell - \mathcal{C}_\ell\right)\left(\hat{C}_{\ell'} - \mathcal{C}_{\ell'}\right) \right\rangle = 2\,\mathrm{Tr}\ \mathbf{C}\mathbf{Z}^\ell \mathbf{C}\mathbf{Z}^{\ell'} \quad \text{(A15)}$$

$$\text{where } \mathbf{Z}^\ell = \tilde{\mathbf{R}}_\ell - \frac{\mathbf{\Delta}^\ell}{2\ell+1}. \quad \text{(A16)}$$

The diagonal part of the cut-induced variance for a 20% azimuthal sky cut is plotted as a band in Figure 2. It may be verified that $\mathrm{CIV}_{\ell\ell'} \neq V_{\ell\ell'}$; expanding expression (A15) shows that the cut-induced variance is equal to the sum of the cut-sky and full-sky cosmic variances minus a unique cross-term.

4. The reconstruction matrices

Let us now turn to specific reconstruction methods. The $\mathbf{R}^\ell$ matrices for the PCL case (*e.g.* Ref. [18]) read

$$\mathbf{R}^\ell_{\mathrm{PCL}} = \sum_{\ell'} (M^{-1})^{\mathrm{PCL}}_{\ell\ell'} \mathbf{\Delta}^{\ell'} / (2\ell'+1), \quad \text{(A17)}$$

$$\text{with } M^{\mathrm{PCL}}_{\ell\ell'} = \mathrm{Tr}\ \mathbf{\Delta}^\ell \tilde{\mathbf{\Delta}}^{\ell'} / (2\ell'+1), \quad \text{(A18)}$$

where the $2\ell'+1$ normalization on each of these expressions is conventional. It may readily be verified that these form unbiased estimates for the full-sky, ensemble-averaged C_ℓ's in an exactly isotropic theory since the covariance matrix $\mathbf{C}$ may be written as

$$\mathbf{C} = \sum_\ell C_\ell \mathbf{\Delta}^\ell. \tag{A19}$$

In this isotropic case one can write the covariance matrix on the cut sky

$$\tilde{\mathbf{C}} = \sum_\ell C_\ell \tilde{\mathbf{\Delta}}^l, \tag{A20}$$

showing that our harmonic-space $\tilde{\mathbf{\Delta}}^\ell$ plays the role of $\mathbf{P}^\ell$ in the notation of Tegmark's pixel-space exposition of the QML estimator [19]. Accordingly, the QML reconstruction matrices are written[7]:

$$\mathbf{R}^\ell_{\rm QML} = \sum_{\ell'} \left(M^{-1}\right)_{\ell\ell'} \tilde{\mathbf{C}}^{-1}\tilde{\mathbf{\Delta}}^{\ell'}\tilde{\mathbf{C}}^{-1} \tag{A21}$$

$$\text{where } M^{\rm QML}_{\ell\ell'} = {\rm Tr}\ \tilde{\mathbf{C}}^{-1}\tilde{\mathbf{\Delta}}^{\ell}\tilde{\mathbf{C}}^{-1}\tilde{\mathbf{\Delta}}^{\ell'}. \tag{A22}$$

In these expressions it is possible to substitute for $\mathbf{C}$ a false covariance matrix $\mathbf{C}^{\rm conc}$ which differs from the true theory matrix used in expressions (A13), (A14) and (A15). This represents the state of affairs when an incorrect assumption is made by an analyst about the isotropy (or some other aspect) of the underlying theory, as simulated in Sec. IV above and Appendix A 5 below. In numerical construction of the QML estimators we assumed a variance on the monopole and dipole of $1000\mu K^2$. This effectively projects out information which is contaminated by cross-talk from the monopole and dipole, and is likely to be over-cautious, but residual foregrounds make it hard to quantify the uncertainty in the WMAP zeroing of these quantities. (See also the discussion in Ref. [19].) We verified the results were not sensitive to the precise variance assumed on $\ell = 0, 1$.

As for the PCL case, the QML estimates are unbiased ($\langle \hat{C}^{\rm QML}_\ell \rangle = C_\ell$) if both $\mathbf{C}$ and $\mathbf{C}^{\rm conc}$ are isotropic. If $\mathbf{C} = \mathbf{C}^{\rm conc}$ they are also optimal in the sense that no unbiased estimator (quadratic or otherwise) can start from the cut-sky $\tilde{a}_{\ell m}$'s and produce C_ℓ estimates with a smaller covariance ellipsoid [19].

However the QML estimator has sometimes been criticized for the dependence of its optimality on the assumed covariance matrix – it appears to rely on the structure of the assumed underlying theory in a way that the PCL estimator does not. [No $\mathbf{C}$ matrices appear in expressions (A17) and (A18).] When the covariance matrix assumed may be incorrect, is it safer to use the PCL estimator? The answer is 'no'; in fact the anisotropy-induced errors in QML estimates are typically smaller than those in PCL estimates. We now explain why this should be the case.

[7] In equation (A21) and below we adopt the convention of assuming the existence of an inverse for the singular matrix $\tilde{\mathbf{C}}$. Practically speaking one can regularize the matrix using an additive numerical noise term, or simply use the pseudo-inverse, since $\tilde{\mathbf{C}}^{-1}$ always appears conjugated by $\mathbf{K}$, the null directions of which lead to the uninvertibility.

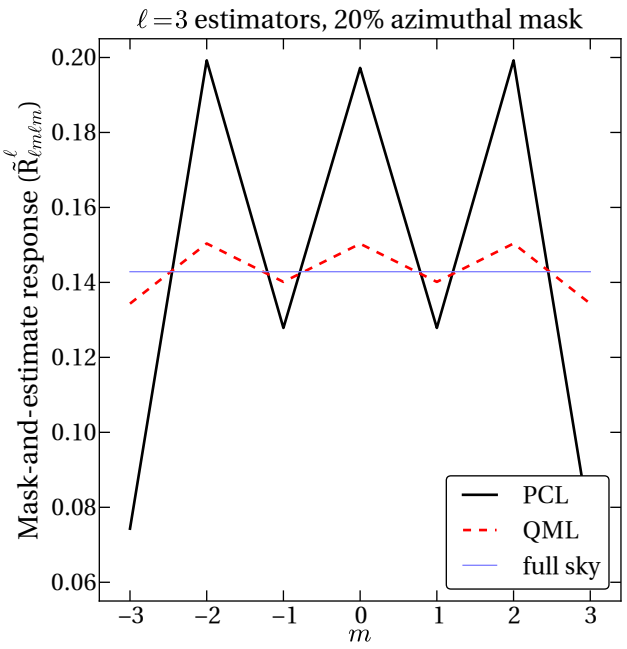

FIG. 5: (Color online.) The weighting which different estimators give to full-sky modes of differing m in estimating the C_ℓ's (here illustrated for $\ell = 3$ with a 20% azimuthal mask). The full sky estimator (thin solid line) by definition weights each m equally; see Eq. (A7). Masking the sky then using the QML estimator (dashed line) comes close to reproducing this weighting, despite the loss of information associated with the first operation. The action of masking the sky then estimating $\ell = 3$ power using the PCL technique (thick solid line) favours $|m| = 2$ and $m = 0$ while downweighting $|m| = 3$ modes, thus giving a less reliable estimate of the full sky power. Similar trends are seen at other ℓ.

5. The relationship between QML and PCL estimators

By examining the relationship between QML and PCL estimators, it is possible to show that QML estimators (derived on the assumption of isotropy) are statistically superior even if the underlying theory breaks isotropy in an unknown way. In outline, the QML estimator can be understood as minimizing the cross-talk from variance in neighbouring ℓ modes. It can only do this by having prior information about the shape of the spectrum in the region of the ℓ estimate under construction. But since the observed sky – regardless of its isotropy – has a $\mathcal{C}_\ell$ power spectrum with a very similar shape to the theoretical

model, the QML estimator is expected to be superior to the PCL estimator under any model compatible with the observed sky.

This argument does rely on the QML estimator not placing undue weight at given ℓ on any particular m-mode. The full-sky estimator by definition gives even weight to each m [see expression (A7)]. To accurately reproduce power spectra, cut-sky estimators must trade off equal weighting of the m modes against down-weighting modes which are particularly contaminated by mask mode-coupling. Derived on the assumption of an isotropic theory, it is not clear whether the QML estimator will do a better or worse job than the PCL estimator in this limited sense. A calculation shows, however, that the QML is superior – it actually weights the full-sky m-modes more evenly than the PCL estimator. This is illustrated in Figure 5, which shows the weight given to each m mode on the full sky under the composite operation of masking-then-estimating. The weights are flatter for the QML estimator (dashed line) than for the PCL estimator (thick solid line). It is actually a fortuitous result of the shape of the concordance power spectrum that this is true; otherwise the reliability of the QML estimator would depend more sensitively on the underlying anisotropic theory.

We now demonstrate the crucial result that the PCL and QML estimators become identical for a flat power spectrum. The covariance matrix is then proportional to the identity, $\mathbf{C} = \alpha\mathbf{I}$, so that the QML reconstruction matrices reduce to

$$\mathbf{R}^{\ell}_{\rm QML} = \alpha^{-2}\,\mathbf{M}^{-1}_{\ell\ell'}\tilde{\mathbf{\Delta}}^{\ell'} = \alpha^{-2}\,\mathbf{M}^{-1}_{\ell\ell'}\mathbf{K}\mathbf{\Delta}^{\ell'}\mathbf{K} \qquad \text{(A23)}$$

$$\mathbf{M}^{\rm QML}_{\ell\ell'} = \alpha^{-2}\,{\rm Tr}\;\tilde{\mathbf{\Delta}}^{\ell}\tilde{\mathbf{\Delta}}^{\ell'} = \alpha^{-2}\,{\rm Tr}\;\mathbf{\Delta}^{\ell}\tilde{\mathbf{\Delta}}^{\ell'} \qquad \text{(A24)}$$

where the final expression for $\mathbf{M}^{\rm QML}$ is obtained by expanding the masked expressions ($\tilde{\mathbf{\Delta}}^{\ell} = \mathbf{K}\mathbf{\Delta}^{\ell}\mathbf{K}$) and using the condition $\mathbf{K}^2 = \mathbf{K}$ obtained in Section A 1.

By rewriting equation (A6) as $\hat{C}_\ell = \mathbf{a}^\dagger\mathbf{K}\mathbf{R}^\ell\mathbf{K}\mathbf{a}$, it follows that the $\mathbf{R}^\ell$ appearing in any statistical expression must arise in the combination $\mathbf{K}\mathbf{R}^\ell\mathbf{K}$. This means that one can, without loss of generality, dispense with the explicit masking $\mathbf{K}$ matrices in expression (A23), again relying on the identity $\mathbf{K}^2 = \mathbf{K}$. Finally, the α^2 factors in equations (A23) and (A24) may be mutually cancelled, since $\mathbf{M}$ appears only in the expression for $\mathbf{R}$. Thus we may write

$$\mathbf{R}^{\ell}_{\rm QML} \sim \mathbf{M}^{-1}_{\ell\ell'}\mathbf{\Delta}^{\ell'} = \mathbf{R}^{\ell}_{\rm PCL} \qquad \text{(A25)}$$

$$\text{where } \mathbf{M}_{\ell\ell'} = {\rm Tr}\;\mathbf{\Delta}^{\ell}\tilde{\mathbf{\Delta}}^{\ell'} \qquad \text{(A26)}$$

where the $\sim$ symbol should be read as 'yields identical estimates to' – *i.e.* it denotes an equivalence relation, not an approximate equality of the matrix elements. To verify this, compare the above with Eqs (A17, A18), noting that the missing factors of $2\ell' + 1$ are conventional normalizations which exactly cancel between the two lines.

This demonstrates that, if $\mathbf{C} = \alpha\mathbf{I}$, QML estimates are identical to PCL estimates. The result does not rely on any assumptions about the Galactic cut being small. However, for a small Galactic cut, the mask operation $\mathbf{K}$ acquires a narrow banded structure at high-ℓ such that each ℓ is effectively coupled only to a finite range of ℓ' from $\ell-\Delta\ell$ to $\ell+\Delta\ell$. Thus, even though the concordance covariance matrix is not proportional to the identity, at high ℓ its relevant, local structure can be adequately approximated as such. This demonstrates the equivalence of the QML and PCL estimators in this regime.

To understand the difference between QML and PCL estimators one can reverse the argument above [*i.e.* one replaces $\mathbf{\Delta}^{\ell'}$ by $\tilde{\mathbf{\Delta}}^{\ell'}$ in equation (A17) then compares with equation (A21), finding the latter simply pre-weights the data].

It follows from all this that the QML estimator can, roughly speaking, be expected to remain superior to the PCL estimator for any theory compatible with our sky. To demonstrate this explicitly, we draw random covariance matrices with power spectrum equal to that of the observed ILC, but taking a random distribution of power between different m modes. Explicit calculations for each of these theories show that the QML biases (A13) and variances (A14) are significantly smaller than their PCL counterparts. A specific illustration is given in Figure 6, where we plot a histogram of the biases on $\hat{C}_5^{\rm PCL}$ (solid line) and $\hat{C}_5^{\rm QML}$ (dashed line) for 200 000 random theories. The PCL estimator has a significantly broader distribution of biases than the QML estimator, showing that the QML technique typically produces more reliable estimates of the full sky power.

In conclusion, for any anisotropic theory which is compatible with our observed sky, estimates for the power spectrum formed using the QML technique (despite being derived assuming isotropy) are expected to be superior to PCL estimates for the same quantity.

Appendix B: Estimators for $C(\theta)$ and $S_{1/2}$

In this section we demonstrate that the pixel-based cut-sky correlation function,

$$\mathcal{C}(\theta)^{\rm cut} \equiv \frac{\int {\rm d}\hat{\mathbf{n}}_1{\rm d}\hat{\mathbf{n}}_2 M(\hat{\mathbf{n}}_1)M(\hat{\mathbf{n}}_2)T(\hat{\mathbf{n}}_1)T(\hat{\mathbf{n}}_2)\,\delta_\theta}{\int {\rm d}\hat{\mathbf{n}}_1\,{\rm d}\hat{\mathbf{n}}_2 M(\hat{\mathbf{n}}_1)M(\hat{\mathbf{n}}_2)\,\delta_\theta}, \quad \text{(B1)}$$

where $\delta_\theta = \delta(\hat{\mathbf{n}}_1\cdot\hat{\mathbf{n}}_2 - \cos\theta)$, is identical to the PCL-based estimator[8]

$$\mathcal{C}(\theta)^{\rm PCL} \equiv \frac{1}{4\pi}\sum_\ell (2\ell+1)\hat{C}_\ell^{\rm PCL}P_\ell(\cos\theta). \qquad \text{(B2)}$$

This has been stated before [17] and is implicit in other works [35, 36], but an explicit demonstration has not, to

[8] We are grateful to A. Challinor for initially drawing our attention to this equivalence.

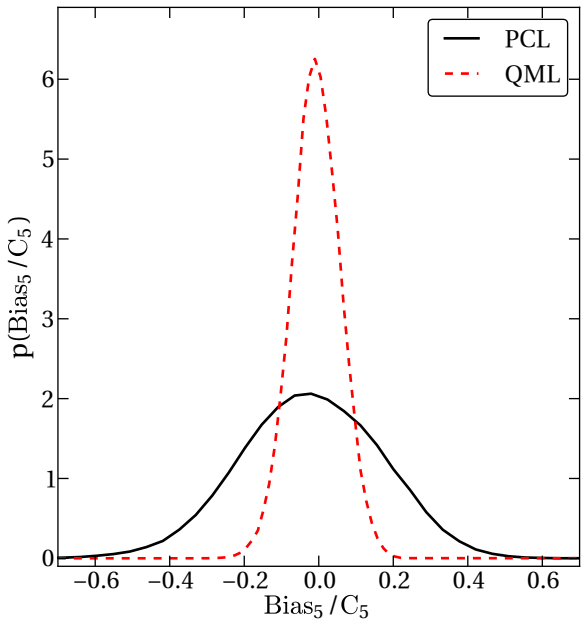

FIG. 6: (Color online.) An example of how the QML estimator typically remains superior to the PCL estimator is given by comparing, for 200 000 random anisotropic theories, the bias on the cut-sky power spectrum estimates (here illustrated for $\ell = 5$ with a 20% azimuthal mask). The width of the curves show that, for a given theory, the QML estimator (dashed curve) is typically significantly less biased than the PCL estimator (solid curve). Similar results hold at other ℓ's and for the extra variance induced by the anisotropy.

our knowledge, appeared in the literature. The argument holds for any weighting function $M(\hat{\mathbf{n}})$: if M takes values other than 0 and 1 the harmonic-space matrix $\mathbf{K}$ [still defined by equation (A2)] is no longer idempotent, but no part of the proof below is affected by such a change.

Equation (B1) may be expressed

$$\mathcal{C}(\theta)^{\rm cut} = \frac{1}{F(\theta)} \sum_\ell \frac{2\ell+1}{4\pi} \tilde{C}_\ell P_\ell(\cos\theta), \qquad \text{(B3)}$$

where $F(\theta)$, equal to the denominator of (B1), is a normalizing function dependent only on $M(\hat{\mathbf{n}})$. Here $\tilde{\mathcal{C}}_\ell$ is the power spectrum of the masked (or weighted) sky,

$$\tilde{C}_\ell = \frac{\tilde{\mathbf{a}}^\dagger \mathbf{\Delta}^\ell \tilde{\mathbf{a}}}{2\ell+1} = M^{\rm PCL}_{\ell\ell'} \hat{C}^{\rm PCL}_{\ell'}, \qquad \text{(B4)}$$

where $\mathbf{M}^{\rm PCL}$ is defined by equation (A18). A power spectrum can be calculated from $\mathcal{C}(\theta)^{\rm cut}$:

$$\hat{C}^{\rm cut}_\ell = 2\pi \int_{-1}^{1} \mathcal{C}(\theta)^{\rm cut} P_\ell(\cos\theta)\, {\rm d}\cos\theta \qquad \text{(B5)}$$

$$= T_{\ell\ell'} \tilde{C}_{\ell'} = T_{\ell\ell'} M^{\rm PCL}_{\ell'\bar{\ell}} \hat{C}^{\rm PCL}_{\bar{\ell}}, \qquad \text{(B6)}$$

where the matrix

$$T_{\ell\ell'} = \frac{2\ell'+1}{2} \int_{-1}^{1} \frac{P_\ell(\cos\theta)\, P_{\ell'}(\cos\theta)}{F(\theta)} {\rm d}\cos\theta \qquad \text{(B7)}$$

depends only on $F(\theta)$ (and hence only on the sky cut, not any aspects of the theory or realization).

If we temporarily consider a theory which is isotropic, we have $\langle \mathcal{C}(\theta)^{\rm cut} \rangle = C(\theta) = \langle \mathcal{C}(\theta)^{\rm PCL} \rangle$ and hence, by linearity, $\langle \hat{C}^{\rm cut}_\ell \rangle = \langle \hat{C}^{\rm PCL}_\ell \rangle$. But then, comparing with expression (B6), the only possibility is that the matrix $T_{\ell\ell'}$ is the inverse of the matrix $M^{\rm PCL}_{\ell\ell'}$ – in other words that

$$\hat{C}^{\rm cut}_\ell = \hat{C}^{\rm PCL}_\ell. \qquad \text{(B8)}$$

We reiterate that neither $\mathbf{T}$ nor $\mathbf{M}^{\rm PCL}$ depend on either the underlying theory nor the particular realization in hand, and therefore this result is independent of isotropy. Finally, one inverts the Legendre transform to gain the desired result,

$$\mathcal{C}(\theta)^{\rm PCL} \equiv \mathcal{C}(\theta)^{\rm cut}, \qquad \text{(B9)}$$

valid for any theory. It follows immediately that $\mathcal{S}^{\rm cut}_{1/2}$ derived from expression (10) must be mathematically equivalent to pixel-based estimates.

The above proof depends on the invertibility of $M^{\rm PCL}_{\ell\ell'}$. It is well known that this matrix is not invertible for all choices of sky-cut (although for all masks considered in the present work we have found it to be well-behaved). However, in any limit where $\det \mathbf{M}^{\rm PCL} \to 0$, one must have $\det \mathbf{T} \to \infty$. According to definition (B7), this will occur if and only if $F(\theta) \to 0$ for some θ – in other words if and only if the cut sky contains, in the limit, no two points separated by certain values of θ. It follows that, whenever the entire correlation function can be recovered from the cut sky, the PCL estimates can be made and the relationship proved above holds.

Aside: $\mathcal{S}^{\rm cut}_{1/2}$ is biased high

We should note in passing that, because $\mathcal{S}_{1/2}$ is quadratic in the $\mathcal{C}_\ell$'s, its expectation value does not follow simply by replacing the $\mathcal{C}_\ell$'s with the C_ℓ's in Eq. (6); rather, the full sky expectation value reads for the concordance theory

$$S_{1/2} \equiv \langle \mathcal{S}_{1/2} \rangle = \sum_{\ell\ell'} s_{\ell\ell'} \left(C_\ell C'_\ell + \frac{2C^2_\ell \delta_{\ell\ell'}}{(2\ell+1)} \right). \qquad \text{(B10)}$$

The second term in Eq. (B10) contributes very significantly to the expectation value, which breaks down term-by-term as $S_{1/2} = (4.9 + 3.7) \times 10^4 = 8.6 \times 10^4\, \mu{\rm K}^4$ for the WMAP5 best fit C_ℓ's. This means that (for instance) the comparison of our full sky with the theory values in Table 1 of Ref. [10] is not strictly appropriate; with the

cosmic variance term included, the observed $\mathcal{S}_{1/2}$ statistic is made to look even more discrepant with the theory. The expectation value of $\mathcal{S}_{1/2}^{\rm cut}$ calculated directly from the cut sky $\hat{C}_\ell$ is also increased, for similar reasons, but by a larger amount corresponding to the larger variance on the cut sky power spectrum estimates:

$$\langle \hat{C}_\ell \hat{C}_{\ell'} \rangle = \langle \mathcal{C}_\ell \mathcal{C}_{\ell'} \rangle - \frac{2C_\ell^2 \delta_{\ell\ell'}}{(2\ell+1)} + V_{\ell\ell'}, \tag{B11}$$

where $V_{\ell\ell'}$ is the variance of the $\hat{C}_\ell$'s, given by expression (A14). This biases cut-sky $\mathcal{S}_{1/2}$ values to be higher than their full-sky counterparts, for instance by $\sim 8300\,\mu{\rm K}^4$ for PCL and $\sim 1100\,\mu{\rm K}^4$ for QML reconstructions with a 20% azimuthal sky cut.

At face value, such biases make it more surprising that the measured $\mathcal{S}_{1/2}$ should be so small and $\mathcal{S}_{1/2}^{\rm cut}$ even smaller. However, the standard deviation of $\mathcal{S}_{1/2}^{\rm cut}$ is very large ($\sim 10^5\,\mu{\rm K}^4$) so that the biases do not have a significant impact on the frequentist significances. Furthermore, and regardless of the magnitude of the biasing, the Monte Carlo techniques used in Refs [10] and [16] are anyway valid (they automatically take the biasing into account). We have therefore included the discussion above only for pedagogical interest.

Appendix C: Designer theory

In Section IV B we used a theory with covariance matrix $\mathbf{C}$ determined by two considerations:

(i) The full sky power spectrum is given by $C_\ell = \mathcal{C}_\ell^{\rm ILC}$, where $\mathcal{C}_\ell^{\rm ILC}$ is the observed power spectrum on the full-sky ILC map;

(ii) The cut sky power spectrum (PCL estimator using a 20% azimuthal cut) is biased, *i.e.* its expectation value does not equal the full sky power; instead we set $\langle \hat{C}_\ell^{\rm PCL} \rangle = C_\ell^{\rm cut}$. To reproduce the causes of our own sky's low $\mathcal{S}_{1/2}^{\rm cut}$, we set $C_\ell^{\rm cut} = 0$ for $\ell = 2, 3, 5$ and 7; at all other ℓ, $C_\ell^{\rm cut} = \mathcal{C}_\ell^{\rm ILC}$.

Both constraints are linear in the full sky covariance matrix [see Eqs. (A9) and (A13)]. While one can construct a matrix $\mathbf{C}$ satisfying these constraints using straight-forward linear algebra, the result is not unique and furthermore it is hard to enforce that $\mathbf{C}$ be positive definite (as it must be to define a valid covariance matrix). Therefore we adopted the package CVXOPT[9] to find a suitable theory $\mathbf{C}$ within the set of positive definite matrices. CVXOPT allows us to find a unique solution by minimizing any convex quadratic form, for which we chose the function Tr $\mathbf{C}^2$. The choice at first appears arbitrary; but schematically, by considering the eigenvalues of $\mathbf{C}$, one can imagine that minimizing Tr $\mathbf{C}^2$ tries to 'equalize power between as many modes as possible'. This in turn is motivated by our attempt to minimize the cosmic variance on $\mathcal{S}_{1/2}^{\rm cut}$, leading to the most peaked likelihood function (and hence best possible likelihood gains over the concordance theory).

These statements can be made somewhat more mathematically concrete, but we did not find a full proof that minimizing Tr $\mathbf{C}^2$ minimizes the cosmic variance of $\mathcal{S}_{1/2}^{\rm cut}$. Instead, Section IV B gave a strict lower bound on the variance of $\mathcal{S}_{1/2}^{\rm cut}$, and stated that our theory comes close to saturating this limit. The remainder of the present appendix explains the origin of such a variance floor.

We start by considering, for simplicity, the full sky $\mathcal{S}_{1/2}$. We also temporarily approximate the $\mathcal{C}_\ell$ likelihood function as Gaussian. Both of these simplifications will be removed in due course; in particular, all of our numerical results use the exact likelihood. The variance of the $\mathcal{S}_{1/2}$ statistic may be written

$$\langle (\mathcal{S}_{1/2})^2 \rangle - \langle \mathcal{S}_{1/2} \rangle^2 = 4\mathbf{c}^\top \mathbf{s}\mathbf{V}^{\rm fs}\mathbf{s}\mathbf{c} + 2\,{\rm Tr}\;\mathbf{s}\mathbf{V}^{\rm fs}\mathbf{s}\mathbf{V}^{\rm fs}, \tag{C1}$$

where $\mathbf{c}$ is a vector composed of the C_ℓ's [as defined by equation (A9)], and $\mathbf{V}^{\rm fs}$ represents the full sky cosmic variance,

$$V_{\ell\ell'}^{\rm fs} = \frac{2\,{\rm Tr}\;\mathbf{C}\mathbf{\Delta}^\ell\mathbf{C}\mathbf{\Delta}^{\ell'}}{(2\ell+1)(2\ell'+1)}. \tag{C2}$$

We wish to minimize Eq. (C1) with respect to $\mathbf{C}$ while keeping C_ℓ constant. Using a standard Lagrange multipler technique, one obtains

$$\sum_{\ell_1\ell_4} s_{\ell_1\ell_2} s_{\ell_3\ell_4} (C_{\ell_1} C_{\ell_4} + V_{\ell_1\ell_4}^{\rm fs}) \mathbf{\Delta}^{\ell_2} \mathbf{C} \mathbf{\Delta}^{\ell_3} = 0, \tag{C3}$$

where $\ell_2 \neq \ell_3$; and

$$C_\ell \mathbf{\Delta}^\ell = \mathbf{\Delta}^\ell \mathbf{C} \mathbf{\Delta}^\ell. \tag{C4}$$

The most obvious solution to the minimization equations (C3) and (C4) is the isotropic one,

$$\mathbf{C} = \sum_\ell C_\ell \mathbf{\Delta}^\ell. \tag{C5}$$

One can verify that the solution (C5) is a minimum (not maximum) of expression (C1). To demonstrate that no other minima exist, consider the only alternative to (C5): namely that $\mathbf{\Delta}^{\ell_2}\mathbf{C}\mathbf{\Delta}^{\ell_3} \neq 0$ and equation (C3) is instead satisfied by making the numerical coefficient vanish. We consider the case where this is true for all ℓ_2, ℓ_3 ($\ell_2 \neq \ell_3$), but the ideas generalize straight-forwardly to the case with only limited numbers of non-zero off-diagonal terms. The most general solution is

$$V_{\ell\ell'} = -C_\ell C_{\ell'} + \sum_i \lambda_i Q^i_{\ell\ell'}, \tag{C6}$$

[9] `http://abel.ee.ucla.edu/cvxopt/`; this package performs convex optimization within a cone. (The space of positive definite matrices is an example of a cone in this sense.)

where the $\mathbf{Q}^i$ are symmetric matrices which satisfy $\mathbf{s}\mathbf{Q}^i\mathbf{s} = 0$ (off-diagonal) and

$$\sum_i \lambda_i Q^i_{\ell\ell} = \frac{2\ell+3}{2\ell+1} C_\ell^2. \tag{C7}$$

The $\mathbf{Q}^i$ may be found numerically using a singular value decomposition technique.

Let us consider whether a physical (positive definite) solution to equations (C6) and (C7) exists. A necessary condition is that

$$|V_{\ell\ell'}| \leq \frac{2C_\ell C_{\ell'}}{\sqrt{(2\ell+1)(2\ell'+1)}}. \tag{C8}$$

This condition is violated by equation (C6) with $\lambda_i = 0$, but can a suitable choice of λ_i remove the violation? There are far fewer $\mathbf{Q}^i$ matrices than degrees of freedom in $\mathbf{V}^{\rm fs}$, so that one would need a numerical coincidence to be able to remove the violation simultaneously at all ℓ. We verified computationally that, for our choice of C_ℓ, this is indeed not possible.

Now when the Gaussian simplification is abandoned, expression (C1) picks up extra terms of the form

$$s_{\ell_1\ell_2}s_{\ell_3\ell_4}\mathrm{Tr}\ \mathbf{C}\boldsymbol{\Delta}^{\ell_1}\mathrm{Tr}\ \mathbf{C}\boldsymbol{\Delta}^{\ell_2}\mathbf{C}\boldsymbol{\Delta}^{\ell_3}\mathbf{C}\boldsymbol{\Delta}^{\ell_4} \text{ and} \tag{C9}$$

$$s_{\ell_1\ell_2}s_{\ell_3\ell_4}\mathrm{Tr}\ \mathbf{C}\boldsymbol{\Delta}^{\ell_1}\mathbf{C}\boldsymbol{\Delta}^{\ell_2}\mathbf{C}\boldsymbol{\Delta}^{\ell_3}\mathbf{C}\boldsymbol{\Delta}^{\ell_4}. \tag{C10}$$

One may verify that with these terms, equation (C5) remains a local minimum of the $\mathcal{S}_{1/2}$ variance. Since there are no other valid local minima in the Gaussian approximation, we should not expect to find new local minima appearing in the non-Gaussian case (the simplification modifies only the skewness, not the width, of the likelihood). A full analytic calculation is prohibitive, but the numerical results quoted in the main paper use the full, non-Gaussian likelihood.

Now let us consider the cut sky case. The de-biased PCL estimates $\hat{\mathbf{c}}$ are related to the power spectrum of the cut sky $\tilde{\mathbf{c}}$ by

$$\hat{\mathbf{c}} = \mathbf{M}^{-1}\tilde{\mathbf{c}} \tag{C11}$$

for the matrix $\mathbf{M}$ defined by equation (A18). (The 'PCL' superscript is dropped for concision.) Because of the exact relation

$$\begin{aligned}\mathcal{S}^{\rm cut}_{1/2} &= \sum_{\ell\ell'} \hat{C}^{\rm PCL}_\ell \hat{C}^{\rm PCL}_{\ell'} s_{\ell\ell'} \\ &= \sum_{\ell_1\ell_2\ell_3\ell_4} M^{-1}_{\ell_1\ell_2} M^{-1}_{\ell_3\ell_4} \tilde{\mathcal{C}}_{\ell_2}\tilde{\mathcal{C}}_{\ell_4} s_{\ell_1\ell_3}\end{aligned} \tag{C12}$$

(see Appendix B) we may define

$$\tilde{\mathbf{s}} = (\mathbf{M}^{-1})^\top \mathbf{s} \mathbf{M}^{-1}, \tag{C13}$$

so that equation (C12) simplifies to

$$\mathcal{S}^{\rm cut}_{1/2} = \sum_{\ell\ell'} \tilde{s}_{\ell\ell'}\tilde{\mathcal{C}}_\ell\tilde{\mathcal{C}}_{\ell'}. \tag{C14}$$

The cut-sky reasoning then follows through exactly as for the full-sky case, except with $\mathbf{s}$ and $\mathbf{C}$ replaced by $\tilde{\mathbf{s}}$ and $\tilde{\mathbf{C}}$ respectively. For the lower bound theory one obtains

$$\tilde{\mathbf{C}} = \sum_{\ell\ell'} M_{\ell\ell'} C^{\rm cut}_{\ell'} \boldsymbol{\Delta}^\ell. \tag{C15}$$

Clearly this ignores the implicit restrictions on $\tilde{\mathbf{C}}$ arising from its status as a cut-sky, rather than full-sky, covariance matrix. (Specifically, a valid $\tilde{\mathbf{C}}$ must live in the cut-sky subspace so that $\tilde{\mathbf{C}} = \mathbf{K}\tilde{\mathbf{C}}\mathbf{K}$.) However the set of all valid $\tilde{\mathbf{C}}$ is, crucially, a subset of the positive-definite matrices which were considered for the full-sky case. Therefore our test theory still gives a lower bound for the set of valid theories.

To actually calculate the lower bound we draw 20 000 sets of $\tilde{a}_{\ell m}$'s according to the covariance matrix (C15) and, for each, calculate the power spectrum $\tilde{\mathcal{C}}_\ell$ and hence $\mathcal{S}^{\rm cut}_{1/2}$ according to equation (C14). Calculating the variance on this random sample leads to the numerical lower limit quoted in Section IV B.

Finally note that, because $\mathbf{M}$ is almost diagonal, our Monte Carlo results are almost equivalent to those obtained by calculating $\mathcal{S}_{1/2}$ in an isotropic theory satisfying

$$\mathbf{C} = \sum_\ell C^{\rm cut}_\ell \boldsymbol{\Delta}^\ell. \tag{C16}$$

This is the justification for our intuitive explanation that the lower bound on the variance of $\mathcal{S}^{\rm cut}_{1/2}$ is given by the variance of $\mathcal{S}_{1/2}$ in an isotropic, full-sky theory with power spectrum equal to the cut-sky power spectrum of the designer theory.

Appendix D: A numerical problem and solution

There is a hidden numerical pitfall in the harmonic-space masking operation as defined by Eq. (A2). The matrix $\mathbf{K}$ is not band-limited, which means that truncating at finite ℓ produces cut-sky vectors $\tilde{\mathbf{a}}$ which retain some information about the data inside the cut. This garbled information is visible in maps as low-amplitude ringing around the edges of the cut. The QML estimator, in particular, is very efficient at regenerating the full sky from this trace of unwanted information.

We investigated two methods of mitigating this problem, both of which generated results in good agreement with pixel space techniques. The first is heuristic, simply smoothing the input (full sky) and output (cut sky) maps to angular scales larger than $180^\circ/\ell_{\rm max}$.

The second, which we adopted for our final results[10], is to use an eigenvector decomposition. For a specified

[10] This method was suggested to us by S. Gratton.

sky fraction f, we calculate $\mathbf{K}$ to finite $\ell_{\rm max}$ and find its eigenvalues and vectors. We then replace the smallest eigenvalues (specifically, a fraction f of the eigenvalues) by zero and all other eigenvalues by one. The final operation is then guaranteed to be idempotent (unlike the ad hoc smoothing approach) and also discards exactly the right fraction of information from the input map. Visually, we found the maps produced looked almost identical to those masked in pixel space. As commented above, we verified that the final estimator results produced from a harmonic-space analysis were closely compatible with those produced from a pixel space analysis. The latter are slow and cumbersome [they cannot take advantage of simplification (A5)], but do not suffer from the band-limitation problem and therefore serve as a useful point of comparison.

Appendix E: Rapid calculation of $L^2_{\rm max}$

In the main text, we discussed the planarity of $\ell = 2$ and $\ell = 3$ power in the observed CMB. This is uncovered [5] by computing the 'angular momentum dispersion'

$$L^2_\ell = \frac{\sum_m m^2 \left|a_{\ell m}\right|^2}{\ell^2 \sum_m \left|a_{\ell m}\right|^2}, \tag{E1}$$

designed to detect 'planarity' of power. By maximizing this quantity across different rotations of the sky, one produces a preferred direction in which the power is most planar. Frequentist anomalies are then reported if the sky-measured maximum values of L^2_ℓ have small P-values according to Monte Carlo simulations of statistically isotropic skies – or if the maximizing directions for two different ℓ's are coincident.

We will show below that calculating L^2_ℓ after rotating the sky by Euler angles[11] $(\phi - \pi/2, -\theta, 0)$ is equivalent to forming the quantity

$$L^2_\ell(\theta,\phi) = \frac{n_i n_j \sum_{mm'} L_{ij}^{mm'} a_{\ell m} a^*_{\ell m'}}{\ell^2 \sum_m \left|a_{\ell m}\right|^2} \tag{E2}$$

where the vector $\mathbf{n} = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$ and the matrix elements $L_{ij}^{mm'}$ are given explicitly below. Hence, for a given set of $a_{\ell m}$'s, the problem of maximizing L^2_ℓ reduces to finding the maximal eigenvector of the 3×3 real symmetric matrix $\sum_{mm'} L_{ij}^{mm'} a_{\ell m} a^*_{\ell m'}$. This algorithm is cheaper by orders of magnitude than numerical maximization methods that appear to have been used to date.

[11] We adopt the convention of [37]; an Euler rotation (α,β,γ) successively rotates the physical sky relative to the fixed, right-handed coordinate system by $-\alpha$, $-\beta$ and $-\gamma$ around the z, x and z axes respectively. The final z rotation would not affect the value of L^2_ℓ, so we fix $\gamma = 0$.

We now prove relation (E2) and give explicit forms for the matrix elements $L_{ij}^{mm'}$. Consider the spin-ℓ function Ψ_ℓ, which may be expanded as

$$\Psi_\ell(\theta,\phi) = \sum_m a_{\ell m} Y_{\ell m}(\theta,\phi). \tag{E3}$$

By considering the angular momentum of this function in the z direction,

$$(\mathbf{e}_z\cdot\hat{\mathbf{J}})\Psi_\ell(\theta,\phi) = -i\partial_\phi \Psi_\ell(\theta,\phi) = \sum_m m a_{\ell m} Y_{\ell m}(\theta,\phi), \tag{E4}$$

where $\hat{\mathbf{J}}$ is the fiducial angular momentum operator and $\mathbf{e}_z$ is the unit vector in the z direction, one may rewrite equation (E1) as

$$L^2_\ell = \frac{\left\langle \Psi_\ell \left| \left(\mathbf{e}_z\cdot\hat{\mathbf{J}}\right)^2 \right| \Psi_\ell \right\rangle}{\ell^2 \langle\Psi_\ell|\Psi_\ell\rangle}, \tag{E5}$$

where the inner product is defined as usual:

$$\langle\Phi|\Psi\rangle = \int \Phi^*(\theta,\phi)\Psi(\theta,\phi)\,\sin\theta\,{\rm d}\theta\,{\rm d}\phi. \tag{E6}$$

To convert expression (E5) for a single L^2_ℓ value into the form (E2), which is claimed to give all possible sky-rotated values, we first need to see that the following two operations are equivalent:

(i) rotating the function by angle α about an axis specified by the vector $\mathbf{r}$, then measuring L^2_ℓ;

(ii) rotating the z coordinate direction by angle $-\alpha$ about $\mathbf{r}$ while keeping the function fixed, *i.e.* replacing $\mathbf{e}_z$ in equation (E5) by the rotated vector $\mathbf{e}_z{}'$.

The equivalence is intuitively clear because the only preferred direction in expression (E5) is given by the $\mathbf{e}_z$ vector. To establish the result more formally, one can use the commutation relations for the angular momentum operators applied to an infinitesimal rotation, and then extend, as usual, to finite rotations by exponentiation.

The result is that

$$L^2_\ell(\theta,\phi) = \frac{\left\langle \Psi_\ell \left| \left(\mathbf{n}\cdot\hat{\mathbf{J}}\right)^2 \right| \Psi_\ell \right\rangle}{\ell^2 \langle\Psi_\ell|\Psi_\ell\rangle} = \frac{n_i n_j \left\langle \Psi_\ell \left| \hat{J}_i\hat{J}_j \right| \Psi_\ell \right\rangle}{\ell^2 \langle\Psi_\ell|\Psi_\ell\rangle}, \tag{E7}$$

with $\mathbf{n}$ defined as above. In harmonic space the operator $\hat{J}_i\hat{J}_j$ forms the matrix elements appearing in Eq. (E2):

$$L_{ij}^{mm'} = \langle Y_{\ell m'}|\hat{J}_i\hat{J}_j|Y_{\ell m}\rangle = \sum_{m''} J_i^{mm''} J_j^{m''m'}, \tag{E8}$$

where the harmonic space angular momentum operators are obtained numerically using the relations

$$\begin{aligned} J_z^{mm'} &= m'\delta_{mm'}; \\ J_x^{mm'} &= \frac{1}{2}\left(J_+^{mm'} + J_-^{mm'}\right) \text{ and} \\ J_y^{mm'} &= \frac{1}{2i}\left(J_+^{mm'} - J_-^{mm'}\right). \end{aligned} \tag{E9}$$

Here the matrix elements of the ladder operators $\hat{J}_\pm$ obey

$$J_\pm^{mm'} = \sqrt{l(l+1) - m(m+1)}\delta_{m\pm1,m'}. \quad \text{(E10)}$$

Finally we note that analytic moments of the statistical distributions may be calculated using the above formalism; however, in practice, the distributions are rather asymmetrical at low ℓ and it is conceptually and computationally easier to use Monte Carlo results – this is extremely fast using the new algorithm, especially since the $L_{ij}^{mm'}$ matrix elements can be precomputed and cached for each ℓ of interest.